

# Theoretical Analysis of Active Contours on Graphs

Christos Sakaridis\*, Kimon Drakopoulos<sup>†</sup>, and Petros Maragos<sup>‡</sup>

\*ETH Zürich, <sup>†</sup>University of Southern California, <sup>‡</sup>National Technical University of Athens

**Abstract**—Active contour models based on partial differential equations have proved successful in image segmentation, yet the study of their geometric formulation on arbitrary geometric graphs is still at an early stage. In this paper, we introduce geometric approximations of gradient and curvature, which are used in the geodesic active contour model. We prove convergence in probability of our gradient approximation to the true gradient value and derive an asymptotic upper bound for the error of this approximation for the class of random geometric graphs. Two different approaches for the approximation of curvature are presented and both are also proved to converge in probability in the case of random geometric graphs. We propose neighborhood-based filtering on graphs to improve the accuracy of the aforementioned approximations and define two variants of Gaussian smoothing on graphs which include normalization in order to adapt to graph non-uniformities. The performance of our active contour framework on graphs is demonstrated in the segmentation of regular images and geographical data defined on arbitrary graphs.

**Index Terms**—Geodesic active contours, graph segmentation, random geometric graphs, image segmentation, object detection.

## I. INTRODUCTION

**E** VOLUTION of curves via active contour models has been applied extensively in computer vision for image segmentation and object detection. In the classical image setting which involves a regular grid of pixels, the discretization of PDEs governing the motion of active contours is well-established and ensures proper convergence of the contour to object boundaries. Recently, active contours have been extended to handle more general input in the form of graphs whose vertices are arbitrarily distributed in a two-dimensional Euclidean space. This arbitrary spatial configuration poses a significant challenge to the discrete approximation of continuous operators that are used in active contours. Applications of segmentation of such graphs span not only image processing, but also geographical information systems and generally any field where data can assume the form of a set of pointwise samples of a real-valued function.

Our work focuses mainly on the theoretical study of fundamental geometric terms in active contours, primarily *gradient* and *curvature*, and the introduction of novel, neighborhood-based approximations of them on arbitrary graphs, which improve upon previous approaches. We analyze the exactness of these approximations and prove convergence to the true values in the limit of large-scale input for the class of *random geometric graphs*. Additionally, we derive an asymptotic bound for the error of our gradient approximation with respect to the number of vertices of the graph. Another important contribution is the usage of neighborhood-based smoothing filtering on graphs as an algorithmic heuristic to reduce the

error of our approximations for smooth inputs. Last, we propose normalized versions of Gaussian filtering on graphs (which is essential for initialization of active contour schemes), suited to handle non-uniform vertex distributions.

The paper is structured as follows. Section II reviews previous work on active contours, graph-based morphology and PDE-based methods on graphs and provides the necessary background on active contour models. In Section III we introduce the basic quantities of our framework and present our geometric approximation of gradient on graphs. We provide conditions for convergence in probability of our approximation in the case of random geometric graphs and analyze the asymptotic behavior of approximation error, which enables an advised selection of parameters for graph construction. In Section IV we give two methods to approximate curvature on graphs, both of which rely on gradient approximation, and state theorems about their convergence in probability for random geometric graphs. Section V is dedicated to defining neighborhood-based smoothing filters on graphs, introducing normalized Gaussian filtering and Gaussian derivative filtering on graphs, and demonstrating their use in smoothing synthetic gradient, curvature or image functions. In Section VI we apply the geodesic active contour algorithm on graphs constructed synthetically, defined from regular images, or containing geographical data, and compare different methods to create the set of vertices and/or edges of these graphs.

## II. BACKGROUND AND RELATED WORK

Active contour models for curve evolution towards image edges originate from “snakes” [1]. These early approaches could not in general handle topological changes of the contour, for instance splitting into two disjoint parts to detect the boundaries of two distinct objects. PDE-based methods using level sets were proposed as an alternative in [2], [3], where the geometric active contour model was initially introduced and subsequently complemented to establish the geodesic active contour (GAC) framework. The former model involves two forces that govern curve motion: a balloon force that expands or shrinks it, and a curvature-dependent force that maintains its smoothness. The latter model adds an extra spring force that attracts the contour towards salient image edges. Both methods embed the active contour as a *level set* of the function  $u$  involved in the PDE that models curve evolution, allowing the use of a numerical scheme of the type proposed in [4].

Graphs have long been connected to image processing, in part through their study in terms of mathematical morphology. The application of morphological transforms on neighborhood graphs was established in [5], while a wide variety of graph structures, algorithms for their construction and early applications in computer vision were surveyed in [6]. The notion of

structuring element in classical morphology was extended to graphs in [7], where the proposed structuring graph enables a generalization of neighborhood functions on a graph beyond the one induced by its set of edges. Morphological operators on graphs have been studied further in [8], where the lattice of the subgraphs of a graph is considered in order to define filters that treat the graph as a whole.

Recently, several works, including [9]–[12], have focused on the construction of PDE-based rather than algebraically defined morphological operators on graphs, which are then used to define active contour models on graphs. All these works are based on the definition of a gradient operator on graphs, however, [9]–[11] work on weighted graphs and define a discrete gradient vector on vertices whose dimensionality is the same as the cardinality of each vertex’s neighborhood, whereas [12] considers unweighted graphs and approximates the continuous gradient at each vertex. In [10], [11], active contours are formulated in a variational framework, while in [12], the gradient approximation is leveraged to translate the aforementioned geodesic active contour segmentation framework to 2D graphs with arbitrary structure and vertex configuration. We follow the latter path and carefully treat the geometric quantities involved in the active contour model, such as gradient and curvature. Our aim is to establish graph-based approximations of these quantities that guarantee proper convergence of the contour to object boundaries and that are exact in the limit of large, dense geometric graphs. In particular, to the best of our knowledge, the asymptotic upper bound for the error of our gradient approximation for random geometric graphs is the first of its kind.

A different class of approaches to graph segmentation which has gained a lot of interest in the image processing community is based on graph cuts. These approaches, in contrast to ours, usually operate on a regular image grid and define weighted edges between image pixels based on certain cues like spatial or appearance proximity, in order to find a cut of minimal cost for the resulting weighted graph. The cost of a cut is normalized in [13] so that balanced partitions are preferred. Approximate solutions to multi-label problems are proposed in [14], guaranteeing constant-factor optimality. A link between geodesic active contours and graph cuts is established in [15], where the graph is constructed so that the cost of the cut corresponds to the contour’s length under the induced anisotropic metric, and this link is extended to the arbitrary graph setting in [12]. Efficient algorithms for watershed-like segmentation that are formulated as graph cuts are introduced in [16], [17]. The power watershed framework of [18] unites and generalizes several graph-based optimization methods for image segmentation by expressing their energies in a common, parametric form.

### III. GRADIENT APPROXIMATION ON GRAPHS

The first term of the active contour evolution model that needs to be approximated is the gradient of the bivariate embedding function. We thus develop a general method for calculating the gradient of a real-valued, bivariate function that is implicitly defined on a continuous domain, although its

values are known only at a sparse, finite set of points, which coincide with the vertices of the graph.

#### A. Main Idea, Notation and Definitions

Compared to the proposals of Drakopoulos and Maragos [12] for gradient approximation, we attempt to incorporate our knowledge about the local spatial configuration of vertices in the approximation. More specifically, we introduce the concept of the *angle* around a vertex which is “occupied” by each of its neighbors and use this concept directly in our novel geometric gradient approximation. Our motivation for this approach comes from the following lemma in bivariate calculus.

*Lemma 1:* The gradient of a differentiable function  $u : \mathbb{R}^2 \rightarrow \mathbb{R}$  at point  $\mathbf{x}$  is

$$\nabla u(\mathbf{x}) = \frac{\int_0^{2\pi} D_\phi u(\mathbf{x}) \mathbf{e}_\phi d\phi}{\pi}, \quad (1)$$

where  $\mathbf{e}_\phi$  is the unit vector in direction  $\phi$  and  $D_\phi u(\mathbf{x})$  is the directional derivative of  $u$  at  $\mathbf{x}$  in this direction, defined by

$$D_\phi u(\mathbf{x}) = \lim_{h \rightarrow 0} \frac{u(\mathbf{x} + h\mathbf{e}_\phi) - u(\mathbf{x})}{h}.$$

Based on Lemma 1, the goal of this section is to approximate the gradient at a vertex of the graph by substituting the integral

$$\mathcal{I} = \int_0^{2\pi} D_\phi u(\mathbf{x}) \mathbf{e}_\phi d\phi \quad (2)$$

with a sum over all the neighbors of the vertex. To this end, we start by introducing several key concepts.

The Euclidean distance between vertices  $v$  and  $w$  of a graph  $\mathcal{G}$  is denoted by  $d(v, w)$  and the unit vector in the direction of the edge  $vw$  starting at  $v$  is denoted by  $\mathbf{e}_{vw}$ . We define  $\phi(w) \in [0, 2\pi)$  as the angle between the vector  $\mathbf{e}_{vw}$  and the horizontal axis, as in Fig. 1. A vertex  $v$  will be alternatively denoted by  $\mathbf{v}$  to declare its position vector. Moreover, we denote by  $\mathcal{N}(v)$  the set of neighbors of  $v$  in  $\mathcal{G}$ , with cardinality  $N(v)$ . For the sake of brevity in notation, this cardinality will be written simply as  $N$ . We write  $\mathcal{N}(v) = \{w_1, w_2, \dots, w_N\}$  so that the angles  $\phi(w_i)$  are in ascending order. Based on this ordering, we define the angle around  $v$  “occupied” by  $w_i$ , which we call *neighbor angle*, as

$$\Delta\phi(w_i) = \begin{cases} \frac{\phi(w_{i+1}) - (\phi(w_N) - 2\pi)}{2} & \text{if } i = 1, \\ \frac{\phi(w_1) + 2\pi - \phi(w_{i-1})}{2} & \text{if } i = N, \\ \frac{\phi(w_{i+1}) - \phi(w_{i-1})}{2} & \text{otherwise.} \end{cases} \quad (3)$$

In a similar fashion, we define the angle corresponding to the bisector between two consecutive neighbors as

$$\omega(w_i) = \begin{cases} \frac{\phi(w_i) + \phi(w_N) - 2\pi}{2} & \text{if } i = 1, \\ \frac{\phi(w_i) + \phi(w_{i-1})}{2} & \text{otherwise.} \end{cases} \quad (4)$$

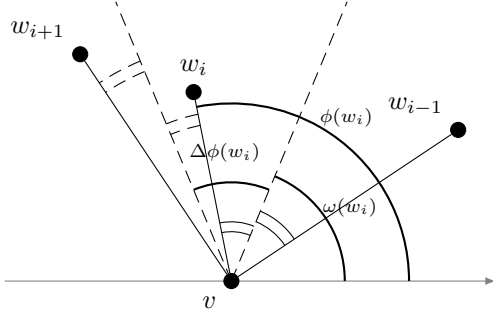


Fig. 1. Angles  $\phi(w_i)$ ,  $\Delta\phi(w_i)$  and  $\omega(w_i)$ .

A visual representation of the neighbor angle is provided in Fig. 1.

Using the above notation, we propose the following formula as the geometric gradient approximation at  $v$ :

$$\nabla u(v) \approx \frac{\sum_{i=1}^N \frac{u(w_i) - u(v)}{d(v, w_i)} \mathbf{e}_{vw_i} \Delta\phi(w_i)}{\pi}. \quad (5)$$

The directional derivative term in (1) is approximated by the difference quotient of the function along each edge. On the other hand, the angle differential is handled through the neighbor angles, which effectively constitute a Voronoi tessellation of the circle around  $v$ , created from its neighbors. The reasoning behind this approach is to use information about the change of  $u$  along each particular direction that comes from the neighbor which is *closest* to this direction.

If we chose not to take the neighbor angles into account, we would place equal importance on all the neighbors of the vertex and we would return to an approximation similar to the weighted sum that was introduced in [12]:

$$\nabla u(v) \approx \frac{\sum_{i=1}^N \frac{u(w_i) - u(v)}{d(v, w_i)} \mathbf{e}_{vw_i}}{N}. \quad (6)$$

### B. Convergence for Random Geometric Graphs

In the following, we will mainly focus on a certain type of graphs, *random geometric graphs* defined below, to study the proposed gradient approximation theoretically.

**Definition 1:** A random geometric graph (RGG)  $\mathcal{G}(n, \rho(n))$  is comprised of a set  $\mathcal{V}$  of vertices and a set  $\mathcal{E}$  of edges. The set  $\mathcal{V}$  consists of  $n$  points distributed uniformly and independently in a bounded region  $D \subset \mathbb{R}^2$ . The set  $\mathcal{E}$  of edges is defined through the radius  $\rho(n)$  of the graph: an edge connects two vertices  $v$  and  $w$  if and only if their distance is at most  $\rho(n)$ , i.e.  $d(v, w) \leq \rho(n)$ .

We show that for this type of graphs, the approximation of (5) converges in probability to the true value of the gradient as the number of vertices increases, under some conditions on the radius, which constrain the density of the graph. Before stating the related theorem, we remind the reader of some definitions for the asymptotic notations which are used in the following analysis.

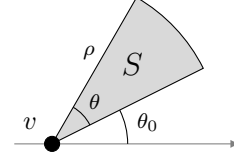


Fig. 2. Sector  $S(v, \rho, \theta_0, \theta)$ .

**Definition 2:** Let  $f$  and  $g$  be two non-negative functions. Then,

$$\begin{aligned} f(n) \in O(g(n)) &\Leftrightarrow \exists k > 0 \exists n_0 \forall n \geq n_0 : f(n) \leq kg(n), \\ f(n) \in \Theta(g(n)) &\Leftrightarrow f(n) \in O(g(n)) \wedge g(n) \in O(f(n)), \\ f(n) \in o(g(n)) &\Leftrightarrow \forall k > 0 \exists n_0 \forall n \geq n_0 : f(n) < kg(n), \\ f(n) \in \omega(g(n)) &\Leftrightarrow \forall k > 0 \exists n_0 \forall n \geq n_0 : f(n) > kg(n). \end{aligned}$$

**Theorem 1:** Let  $u : \mathbb{R}^2 \rightarrow \mathbb{R}$  be a differentiable function and  $\mathcal{G}(n, \rho(n))$  an RGG embedded in  $[0, 1]^2$ , with  $\rho(n) \in \omega(n^{-1/2}) \cap o(1)$ . For a vertex  $v$  of  $\mathcal{G}$ , the approximation of (5) converges in probability to  $\nabla u(v)$ .

*Proof:* It suffices to prove that the sum

$$S = \sum_{i=1}^N \frac{u(w_i) - u(v)}{d(v, w_i)} \mathbf{e}_{vw_i} \Delta\phi(w_i) \quad (7)$$

in (5) converges in probability to the integral (2). Firstly, we show that the norm  $\Lambda_n = \sup_{w \in \mathcal{N}(v)} \{\Delta\phi(w)\}$  of the partition of  $[0, 2\pi]$  induced by the neighbor angles converges in probability to 0 in the limit of large graphs.

Let  $S(v, \rho, \theta_0, \theta)$  be a circular sector centered at  $v$ , with radius  $\rho$ , occupying an angle  $\theta > 0$  and whose rightmost radius is in the direction  $\theta_0$  (see Fig. 2). For every vertex  $z_i, i = 1, \dots, n$  other than  $v$ , we can define a Bernoulli random variable indicating whether this vertex is inside  $S$ :  $Z_i \sim \text{Bern}(\rho^2 \theta / 2)$ . The sum of these variables follows a binomial distribution:

$$Z = \sum_{\substack{i=1 \\ z_i \neq v}}^n Z_i \sim \text{Bin}\left(n-1, \frac{\rho^2(n) \theta}{2}\right).$$

Therefore, the probability that  $S$  is empty of vertices other than  $v$  is  $P(Z = 0) = (1 - \rho^2(n) \theta / 2)^{n-1}$ , which converges to 0, because  $\rho(n) \in \omega(n^{-1/2})$ . Let us consider a neighbor  $w$  of  $v$  and the event

$$\begin{aligned} A &= \exists y, z \in \mathcal{V} \setminus \{v, w\} : y \neq z \\ &\wedge y \in S\left(v, \rho, \phi(w), \frac{\theta}{2}\right) \\ &\wedge z \in S\left(v, \rho, \phi(w) - \frac{\theta}{2}, \frac{\theta}{2}\right), \end{aligned}$$

i.e. there is another neighbor of  $v$  “closer” than  $\theta/2$  on each side of  $w$ . According to the above analysis, it is straightforward that  $\lim_{n \rightarrow +\infty} P(A) = 1$ . Moreover,  $A$  implies  $\Delta\phi(w) \leq \theta$ . Thus, it holds that  $\lim_{n \rightarrow +\infty} P(\Delta\phi(w) \leq \theta) = 1$  and consequently

$$\lim_{n \rightarrow +\infty} P(\Lambda_n \leq \theta) = 1 \quad \forall \theta > 0, \quad (8)$$

which concludes the first part of the proof.

Secondly, we show that the sum

$$\mathcal{S}_1 = \sum_{i=1}^N D_{\phi(w_i)} u(v) \mathbf{e}_{vw_i} \Delta\phi(w_i)$$

converges in probability to  $\mathcal{I}$ .  $\mathcal{S}_1$  constitutes a Riemann sum of  $\mathbf{f}(\phi) = D_{\phi} u(v) \mathbf{e}_{\phi}$  over  $[0, 2\pi]$  with respect to the partition induced by the neighbor angles. Therefore, (8) directly implies that  $\mathcal{S}_1$  converges in probability to  $\mathcal{I}$ .

The third step of the proof is to show that the difference  $\mathcal{S} - \mathcal{S}_1$  converges in probability to 0. Using the triangle inequality, we obtain

$$\|\mathcal{S} - \mathcal{S}_1\| \leq \sum_{i=1}^N \left| \frac{u(w_i) - u(v)}{d(v, w_i)} - D_{\phi(w_i)} u(v) \right| \Delta\phi(w_i).$$

For every  $w \in \mathcal{N}(v)$ , the first order Taylor approximation of  $u$  at  $v$  in the direction  $\phi(w)$  yields

$$|u(w) - u(v) - d(v, w) D_{\phi(w)} u(v)| \in O(d^2(v, w)).$$

We divide both sides with  $d(v, w)$  and use the fact that  $0 \leq d(v, w) \leq \rho(n)$  to arrive at

$$\left| \frac{u(w) - u(v)}{d(v, w)} - D_{\phi(w)} u(v) \right| \in O(\rho(n)).$$

The last result holds for every neighbor of  $v$ , so we can substitute each term of the sum to get

$$\|\mathcal{S} - \mathcal{S}_1\| \in \sum_{i=1}^N O(\rho(n)) \Delta\phi(w_i) = O(\rho(n)) \sum_{i=1}^N \Delta\phi(w_i).$$

The sum of neighbor angles over all neighbors is constant and equals  $2\pi$ , which in turn implies that

$$\|\mathcal{S} - \mathcal{S}_1\| \in 2\pi O(\rho(n)) = O(\rho(n)). \quad (9)$$

If we further make use of the fact that  $\rho(n) \in o(1)$ , we get that  $\|\mathcal{S} - \mathcal{S}_1\| \in o(1)$ . Consequently,  $\mathcal{S} - \mathcal{S}_1$  converges to 0 almost surely and hence in probability as well.

Finally, we combine the results from the second and third step to conclude the proof. For arbitrary  $\epsilon > 0$ , we use the triangle inequality to get

$$\begin{aligned} \mathbb{P}(\|\mathcal{S} - \mathcal{I}\| > \epsilon) &\leq \mathbb{P}(\|\mathcal{S} - \mathcal{S}_1\| + \|\mathcal{S}_1 - \mathcal{I}\| > \epsilon) \\ &\leq \mathbb{P}(\|\mathcal{S} - \mathcal{S}_1\| > \epsilon) + \mathbb{P}(\|\mathcal{S}_1 - \mathcal{I}\| > \epsilon). \end{aligned}$$

The above inequalities hold in the limit as well:

$$\begin{aligned} &\lim_{n \rightarrow +\infty} \mathbb{P}(\|\mathcal{S} - \mathcal{I}\| > \epsilon) \\ &\leq \lim_{n \rightarrow +\infty} \mathbb{P}(\|\mathcal{S} - \mathcal{S}_1\| > \epsilon) + \mathbb{P}(\|\mathcal{S}_1 - \mathcal{I}\| > \epsilon) = 0, \end{aligned}$$

where the last equality is due to convergence in probability of  $\mathcal{S} - \mathcal{S}_1$  to 0 and of  $\mathcal{S}_1$  to  $\mathcal{I}$ . ■

### C. Asymptotic Analysis of Approximation Error

Going one step further, we decompose the error introduced by the geometric gradient approximation in order to obtain a bound on the *rate* of convergence to the true gradient as the size of the RGG grows large. Since the framework in RGGs is stochastic, our results involve *expectations* for the various quantities.

Let us denote the error in approximating  $\mathcal{I}$  with  $\mathcal{S}$  by  $\mathcal{E} = \mathcal{S} - \mathcal{I}$ . Comparing the two expressions, we deduce that the approximation in  $\mathcal{S}$  is threefold:

- 1) Directional derivatives along edges are approximated with difference quotients.
- 2) The approximate value for the directional derivative along each edge is used as a constant estimate for all the directions “falling into” the respective neighbor angle.
- 3) The unit vector in the direction of each edge is also used for all the directions corresponding to the respective neighbor angle.

To isolate the above sources of error, we construct intermediate expressions between  $\mathcal{S}$  and  $\mathcal{I}$  and bound the magnitude of the resulting differences.

**Theorem 2:** Let  $u : \mathbb{R}^2 \rightarrow \mathbb{R}$  be a differentiable function and  $\mathcal{G}(n, \rho(n))$  an RGG embedded in  $[0, 1]^2$ , with  $\rho(n) \in \omega(n^{-1/2}) \cap o(1)$ . For every vertex  $v$  of  $\mathcal{G}$ , it holds that

$$\mathbb{E}[\|\mathcal{E}\|] \in O\left(\rho(n) + \frac{1}{n \rho^2(n)}\right). \quad (10)$$

*Proof:* The first intermediate expression is  $\mathcal{S}_1$ , corresponding to the first part of the error

$$\begin{aligned} \mathcal{E}_1 &= \mathcal{S} - \mathcal{S}_1 \\ &= \sum_{i=1}^N \left( \frac{u(w_i) - u(v)}{d(v, w_i)} - D_{\phi(w_i)} u(v) \right) \mathbf{e}_{vw_i} \Delta\phi(w_i). \end{aligned}$$

We have shown in the proof of Theorem 1 that  $\|\mathcal{E}_1\|$  is bounded asymptotically by the radius of the graph (c.f. (9)), as the difference quotients which are used in  $\mathcal{S}$  are first order approximations of the corresponding directional derivatives:

$$\|\mathcal{E}_1\| \in O(\rho(n)). \quad (11)$$

We define the second intermediate expression as

$$\mathcal{S}_2 = \sum_{i=1}^N \int_{\omega(w_i)}^{\omega(w_i) + \Delta\phi(w_i)} D_{\phi} u(v) \mathbf{e}_{vw_i} d\phi.$$

If we denote the direction of  $\nabla u(v)$  by  $\theta$ , we can write  $D_{\phi} u(v) = \|\nabla u(v)\| \cos(\theta - \phi)$  and the second part of the error can be expressed as

$$\begin{aligned} \mathcal{E}_2 &= \mathcal{S}_1 - \mathcal{S}_2 \\ &= \|\nabla u(v)\| \sum_{i=1}^N \mathbf{e}_{vw_i} \int_{\omega(w_i)}^{\omega(w_i) + \Delta\phi(w_i)} (\cos(\theta - \phi(w_i)) - \cos(\theta - \phi)) d\phi. \end{aligned}$$

After performing some calculations, we obtain

$$\begin{aligned} & \int_{\omega(w_i)}^{\omega(w_i) + \Delta\phi(w_i)} \cos(\theta - \phi) d\phi \\ &= 2 \cos\left(\theta - \omega(w_i) - \frac{\Delta\phi(w_i)}{2}\right) \sin\left(\frac{\Delta\phi(w_i)}{2}\right). \end{aligned}$$

The first factor on the right-hand side can be expanded using the identity  $\cos(a - b) = \cos(a)\cos(b) + \sin(a)\sin(b)$  as

$$\begin{aligned} & \cos\left(\theta - \omega(w_i) - \frac{\Delta\phi(w_i)}{2}\right) \\ &= \cos(\theta - \phi(w_i)) \cos\left(\omega(w_i) + \frac{\Delta\phi(w_i)}{2} - \phi(w_i)\right) \\ &+ \sin(\theta - \phi(w_i)) \sin\left(\omega(w_i) + \frac{\Delta\phi(w_i)}{2} - \phi(w_i)\right). \end{aligned}$$

Furthermore, if we apply the definitions of angles  $\omega(w_i)$  and  $\Delta\phi(w_i)$ , we get the following bound:

$$\begin{aligned} & \left| \omega(w_i) + \frac{\Delta\phi(w_i)}{2} - \phi(w_i) \right| \leq \frac{\Delta\phi(w_i)}{2} \\ & \Rightarrow \omega(w_i) + \frac{\Delta\phi(w_i)}{2} - \phi(w_i) \in O(\Delta\phi(w_i)) \end{aligned}$$

We combine the Taylor expansions of the sine and cosine functions around 0

$$\cos(a) = 1 + O(a^2) \text{ and } \sin(a) = a + O(a^3) = O(a)$$

with the above bound into

$$\begin{aligned} & 2 \cos\left(\theta - \omega(w_i) - \frac{\Delta\phi(w_i)}{2}\right) \sin\left(\frac{\Delta\phi(w_i)}{2}\right) \\ &= 2 \left( \cos(\theta - \phi(w_i)) \left(1 + O(\Delta\phi(w_i)^2)\right) \right. \\ &+ \left. \sin(\theta - \phi(w_i)) O(\Delta\phi(w_i)) \right) \left( \frac{\Delta\phi(w_i)}{2} + O(\Delta\phi(w_i)^3) \right) \\ &= \cos(\theta - \phi(w_i)) \Delta\phi(w_i) + \cos(\theta - \phi(w_i)) O(\Delta\phi(w_i)^3) \\ &+ \sin(\theta - \phi(w_i)) O(\Delta\phi(w_i)^2). \end{aligned}$$

We substitute the above expression into  $\mathcal{E}_2$  and use the triangle inequality and the fact that  $\|\mathbf{e}_{vw_i}\| = 1 \forall i \in \{1, \dots, N\}$  to obtain

$$\begin{aligned} \|\mathcal{E}_2\| &\in \|\nabla u(v)\| \sum_{i=1}^N \left| \cos(\theta - \phi(w_i)) O(\Delta\phi(w_i)^3) \right. \\ &+ \left. \sin(\theta - \phi(w_i)) O(\Delta\phi(w_i)^2) \right|. \end{aligned}$$

Additionally, the absolute values of the sine and the cosine in the last result are bounded from above by 1. Thus, the second part of the error is bounded by

$$\begin{aligned} \|\mathcal{E}_2\| &\in \|\nabla u(v)\| \sum_{i=1}^N O(\Delta\phi(w_i)^2) \\ &= \|\nabla u(v)\| O\left(\sum_{i=1}^N \Delta\phi(w_i)^2\right). \end{aligned} \quad (12)$$

The next step is to take the expectation for both sides of (12). Expectation preserves inequalities and it is straightforward that  $f \in O(g)$  implies  $\mathbb{E}[f] \in O(\mathbb{E}[g])$  for two sequences of random variables  $f$  and  $g$ . As a result, it holds that

$$\mathbb{E}[\|\mathcal{E}_2\|] \in \|\nabla u(v)\| O\left(\mathbb{E}\left[\sum_{i=1}^N \Delta\phi(w_i)^2\right]\right).$$

Since  $N$  is itself a random variable, we employ the law of total expectation and take advantage of the fact that all  $\Delta\phi(w_i)$  are identically distributed to write

$$\begin{aligned} & \mathbb{E}\left[\sum_{i=1}^N \Delta\phi(w_i)^2\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\sum_{i=1}^N \Delta\phi(w_i)^2 \middle| N\right]\right] = \mathbb{E}\left[\sum_{i=1}^N \mathbb{E}\left[\Delta\phi(w_i)^2 \middle| N\right]\right] \\ &= \mathbb{E}\left[N \mathbb{E}\left[\Delta\phi(w)^2 \middle| N\right]\right] \text{ for some } w \in \mathcal{N}(v). \end{aligned}$$

Let us focus on the term  $\mathbb{E}\left[\Delta\phi(w)^2 \middle| N\right]$ . In order to calculate this conditional expectation, we examine the distribution of the random variable  $\Delta\phi(w)$ . The probability that  $\Delta\phi(w) \leq x$  is equal to the probability that at least two neighbors of  $v$  other than  $w$  fall inside the  $2x$  radial interval. Taking into account all possible combinations, it follows that for  $N \geq 3$

$$\begin{aligned} \mathbb{P}(\Delta\phi(w) \leq \pi x) &= \sum_{k=2}^{N-1} \binom{N-1}{k} x^k (1-x)^{N-1-k} \\ &= 1 - (N-1)x(1-x)^{N-2} - (1-x)^{N-1}, \end{aligned}$$

for  $x \in [0, 1]$ . The corresponding PDF of the random variable  $\Delta\phi(w)/\pi$  is

$$(N-1)(N-2)x(1-x)^{N-3}, \quad x \in [0, 1]$$

and therefore  $\Delta\phi(w)/\pi$  follows a Beta distribution with parameters  $\alpha = 2$  and  $\beta = N - 2$ . We use the formulas for the mean and variance of a Beta distribution with known parameters to write

$$\begin{aligned} \mathbb{E}\left[\left(\frac{\Delta\phi(w)}{\pi}\right)^2 \middle| N\right] &= \text{Var}\left[\frac{\Delta\phi(w)}{\pi}\right] + \left(\mathbb{E}\left[\frac{\Delta\phi(w)}{\pi}\right]\right)^2 \\ &= \frac{2(N-2)}{N^2(N+1)} + \frac{4}{N^2} = \frac{6}{N(N+1)}. \end{aligned}$$

Due to linearity of expectation, the conditional expectation we are after is

$$\mathbb{E}\left[\Delta\phi(w)^2 \middle| N\right] = \frac{6\pi^2}{N(N+1)}. \quad (13)$$

Using (13), we obtain

$$\begin{aligned} \mathbb{E}[\|\mathcal{E}_2\|] &\in \|\nabla u(v)\| O\left(\mathbb{E}\left[N \frac{6\pi^2}{N(N+1)}\right]\right) \\ &= \|\nabla u(v)\| O\left(\mathbb{E}\left[\frac{1}{N+1}\right]\right). \end{aligned}$$

We compute the expectation  $E[1/(N+1)]$  using the binomial distribution of the number of neighbors of  $v$ ,  $N \sim \text{Bin}(n-1, \pi\rho^2)$ . The definition of this expectation is

$$\begin{aligned} E\left[\frac{1}{N+1}\right] &= \sum_{k=0}^{n-1} \frac{1}{k+1} \binom{n-1}{k} (\pi\rho^2)^k (1-\pi\rho^2)^{n-1-k} \\ &= \frac{1}{n\pi\rho^2} \sum_{k=0}^{n-1} \binom{n}{k+1} (\pi\rho^2)^{k+1} (1-\pi\rho^2)^{n-(k+1)} \\ &= \frac{1}{n\pi\rho^2} \sum_{k=1}^n \binom{n}{k} (\pi\rho^2)^k (1-\pi\rho^2)^{n-k} \\ &= \frac{1}{n\pi\rho^2} \left(1 - \binom{n}{0} (\pi\rho^2)^0 (1-\pi\rho^2)^n\right) \\ &= \frac{1}{n\pi\rho^2} \left(1 - (1-\pi\rho^2)^n\right). \end{aligned}$$

Since  $\rho(n) \in \omega(n^{-1/2}) \cap o(1)$ , it follows that

$$(1 - (1 - \pi\rho^2(n))^n) \in \Theta(1).$$

Consequently, the expectation of the second part of the error is bounded through

$$E[\|\mathcal{E}_2\|] \in \|\nabla u(v)\| O\left(\frac{1}{n\rho^2(n)}\right). \quad (14)$$

Finally, the third part of the approximation error is

$$\begin{aligned} \mathcal{E}_3 &= \mathcal{S}_2 - \mathcal{I} \\ &= \|\nabla u(v)\| \sum_{i=1}^N \int_{\omega(w_i)}^{\omega(w_i) + \Delta\phi(w_i)} \cos(\theta - \phi) (\mathbf{e}_{vw_i} - \mathbf{e}_\phi) d\phi. \end{aligned}$$

For the  $i$ -th term of the above sum, it holds that  $|\phi - \phi(w_i)| \leq \Delta\phi(w_i)$ , which further implies that

$$\begin{aligned} \|\mathbf{e}_{vw_i} - \mathbf{e}_\phi\| &\leq \|\mathbf{e}_{vw_i} - \mathbf{e}_{\phi(w_i) + \Delta\phi(w_i)}\| \\ &= 2 \sin\left(\frac{\Delta\phi(w_i)}{2}\right). \end{aligned}$$

Thus, the magnitude of  $\mathcal{E}_3$  can be bounded using the triangle inequality as follows:

$$\begin{aligned} \|\mathcal{E}_3\| &\leq \|\nabla u(v)\| \sum_{i=1}^N \int_{\omega(w_i)}^{\omega(w_i) + \Delta\phi(w_i)} \|\mathbf{e}_{vw_i} - \mathbf{e}_\phi\| d\phi \\ &\leq \|\nabla u(v)\| \sum_{i=1}^N \int_{\omega(w_i)}^{\omega(w_i) + \Delta\phi(w_i)} 2 \sin\left(\frac{\Delta\phi(w_i)}{2}\right) d\phi \\ &= \|\nabla u(v)\| \sum_{i=1}^N 2\Delta\phi(w_i) \sin\left(\frac{\Delta\phi(w_i)}{2}\right) \\ &\in \|\nabla u(v)\| \sum_{i=1}^N O\left(\Delta\phi(w_i)^2\right). \end{aligned}$$

The last bound is the same as the one that has been derived in (12) for  $\|\mathcal{E}_2\|$ , which yields:

$$E[\|\mathcal{E}_3\|] \in \|\nabla u(v)\| O\left(\frac{1}{n\rho^2(n)}\right). \quad (15)$$

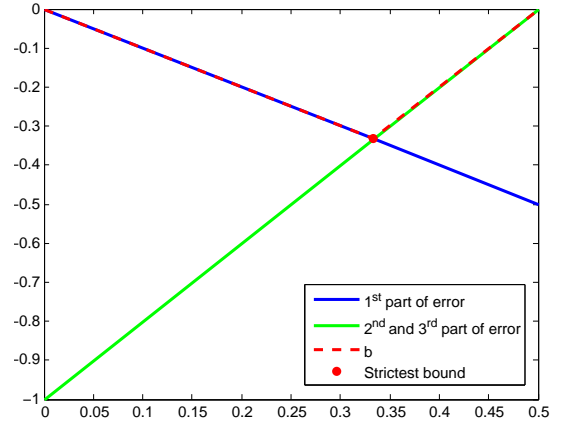


Fig. 3. Variation of exponent  $b$  of the asymptotic bound for gradient approximation error, with respect to exponent  $a$  of the radius. Smaller values of  $b$  mean stricter error bounds. The exponents of the individual parts of the error are also presented.

The total error is  $\mathcal{E} = \mathcal{E}_1 + \mathcal{E}_2 + \mathcal{E}_3$  and due to the triangle inequality and the fact that expectation preserves inequalities, it follows that

$$E[\|\mathcal{E}\|] \leq E[\|\mathcal{E}_1\|] + E[\|\mathcal{E}_2\|] + E[\|\mathcal{E}_3\|].$$

Based on the asymptotic bounds in (11), (14) and (15) and the sum property of the  $O$  symbol, we derive the bound of the total error

$$E[\|\mathcal{E}\|] \in O\left(\rho(n) + \frac{1}{n\rho^2(n)}\right).$$

The radius of the graph is effectively the factor that determines the strictness of this bound. To provide better intuition, we study the case when  $\rho(n) \in \Theta(n^{-a})$ ,  $a \in (0, 1/2)$ . Substituting in (10), we obtain

$$E[\|\mathcal{E}\|] \in O(n^b), \quad b = \begin{cases} -a, & a \in (0, \frac{1}{3}) \\ -1 + 2a, & a \in (\frac{1}{3}, \frac{1}{2}) \end{cases} \quad (16)$$

We visualize this expression for the error bound in Fig. 3. The strictest upper bound is  $O(n^{-1/3})$ , it is achieved for  $a = 1/3$  and it constitutes a tradeoff between minimizing the first error term, which calls for small radii, and the two other terms, which requires more neighbors and consequently larger radii.

From a practical point of view, (14) and (15) indicate that the error in the approximation of (5) increases as the magnitude of the true gradient grows large, i.e. when the function exhibits abrupt variations. This does not pose a problem for the calculation of gradient direction, since the latter does not depend on the range of the function's variation around the examined vertex. Utilizing all incident edges in the weighted sum of (5) ensures that all available information in the neighborhood of the vertex is used to estimate which direction the gradient points to, as emphasized in [12]. However, the estimated gradient *magnitude* with our approximation is prone to greater error, as it depends on the range of the function's variation. The use of difference quotients in (5) accentuates this effect for dense graphs, where distances between neighboring vertices that appear in the denominator

of the quotients approach zero. To circumvent this issue in practice, we adopt the approximation of [12] for gradient magnitude, namely the maximum absolute difference of values of the function along edges that are incident on  $v$ :

$$\|\nabla u(v)\| \approx \max_{w \in \mathcal{N}(v)} \{|u(w) - u(v)|\}. \quad (17)$$

#### IV. CURVATURE APPROXIMATION ON GRAPHS

After having devised an approximation scheme for the gradient of an embedding function, the next step is to use this scheme for estimating the curvature of the level sets of this function. The difference from the gradient case is that the input gradient values for curvature approximation are already approximate themselves, i.e. a *cascaded* approximation is attempted. Therefore, the error in curvature approximation on a graph is expected to accumulate compared to gradient approximation error on the same graph, since the estimated curvature at a vertex inherits the error of the estimated gradients at its neighboring vertices.

##### A. Geometric Approximation

In this type of curvature approximation, we follow an approach similar to [12]. More specifically, we exploit the expression of curvature as the divergence of the unit gradient field  $\mathbf{F} = \nabla u / \|\nabla u\|$  of the embedding function  $u$ :

$$\kappa(v) = \text{div } \mathbf{F}(v), \quad \nabla u(v) \neq \mathbf{0}. \quad (18)$$

An integral definition of divergence as

$$\text{div } \mathbf{F}(v) = \lim_{S \rightarrow \{v\}} \frac{\oint_{\Gamma(S)} \mathbf{F} \cdot \mathbf{n} \, d\ell}{|S|} \quad (19)$$

can then be used as a basis for geometric approximations of curvature, where  $S$  is a region with area  $|S|$  and boundary  $\Gamma(S)$  and  $\mathbf{n}$  is the outward unit normal to this boundary. In [12], the integral in (19) is approximated using a polygonal region to form a finite sum over the neighbors of the vertex (as shown in [12, Fig. 8]). However, certain arrangements of the neighbors of the examined vertex, which are shown in Fig. 4, can lead to regions with ill-defined area, boundary and normals.

To tackle these problems, we employ again the *neighbor angles* that were introduced in Section III, in order to define the region  $S$  in (19) in a more compact and principled fashion. For vertex  $v$ ,  $S(v)$  is formed as a union of circular sectors, each of them corresponding to a neighbor of  $v$ , as we show in Fig. 4(c). More formally, for each neighbor  $w$  of  $v$ , the respective circular sector is  $S(v, d(v, w), \omega(w), \Delta\phi(w))$ . The area of  $S(v)$  can then be expressed as

$$|S(v)| = \sum_{i=1}^N \frac{\Delta\phi(w_i)}{2} d^2(v, w_i). \quad (20)$$

The challenge imposed by our construction of  $S(v)$  is the choice of suitable values for  $\mathbf{F}$  along the boundary of this region, given only its values at the locations of neighbors of  $v$ . The resulting boundary consists of arcs, each of which contains a neighbor of  $v$ , and line segments which connect these

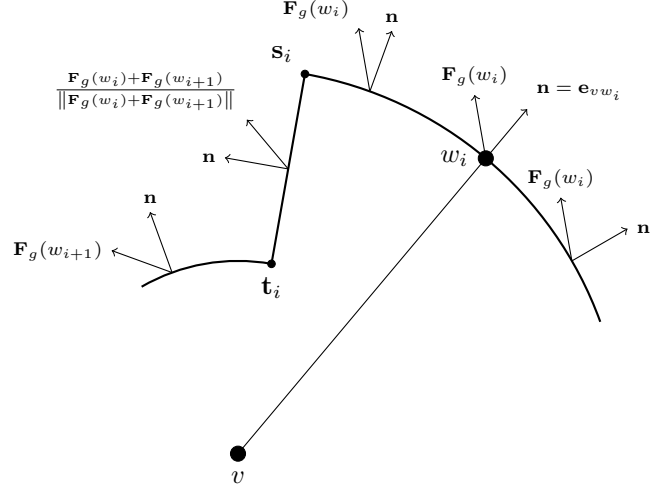


Fig. 5. The picture of values of  $\mathbf{F}_g$  at a part of the boundary of  $S(v)$  which corresponds to neighbor  $w_i$  of  $v$ .

arcs. We fix the value of  $\mathbf{F}$  along each arc at the geometric approximation computed for the corresponding neighbor  $w$  using (5),  $\mathbf{F}_g(w)$ . Moreover, for every line segment, we use the normalized mean of the approximate values of  $\mathbf{F}$  along the two neighboring arcs. The concept is again to use information from the closest vertex, which should be more reliable. We visualize the described configuration in Fig. 5.

Using the above approximations, we substitute the line integral in (19) with a sum of simple line integrals over single arcs and line segments, which have closed analytical forms. If we denote the integral over the arc  $C_a(w_i)$  containing neighbor  $w_i$  by  $I_a(w_i)$  and the integral over the line segment  $C_l(w_i)$  that connects the arcs  $C_a(w_i)$  and  $C_a(w_{i+1})$  by  $I_l(w_i)$ , we obtain

$$\begin{aligned} I_a(w_i) &= \int_{C_a(w_i)} \mathbf{F}_g(w_i) \cdot \mathbf{n} \, d\ell \\ &= d(v, w_i) \mathbf{F}_g(w_i) \cdot (\sin(\omega(w_{i+1})) - \sin(\omega(w_i)), \\ &\quad \cos(\omega(w_i)) - \cos(\omega(w_{i+1}))) \end{aligned} \quad (21)$$

and

$$\begin{aligned} I_l(w_i) &= \int_{C_l(w_i)} \frac{\mathbf{F}_g(w_i) + \mathbf{F}_g(w_{i+1})}{\|\mathbf{F}_g(w_i) + \mathbf{F}_g(w_{i+1})\|} \cdot \mathbf{n} \, d\ell \\ &= (d(v, w_{i+1}) - d(v, w_i)) \frac{\mathbf{F}_g(w_i) + \mathbf{F}_g(w_{i+1})}{\|\mathbf{F}_g(w_i) + \mathbf{F}_g(w_{i+1})\|} \\ &\quad \cdot (\sin(\omega(w_{i+1})), -\cos(\omega(w_{i+1}))). \end{aligned} \quad (22)$$

The proposed geometric approximation of curvature is given by

$$\kappa(v) \approx \frac{\sum_{i=1}^N I_a(w_i) + I_l(w_i)}{|S(v)|}. \quad (23)$$

For RGGs, this approximation is exact in the limit of large graphs like in the gradient approximation case, although the conditions are now stronger.

**Theorem 3:** Let  $\mathcal{G}(n, \rho(n))$  be an RGG embedded in  $[0, 1]^2$ , with  $\rho(n) \in \omega(n^{-1/2}) \cap o(1)$  and  $v$  a vertex of  $\mathcal{G}$ . If  $u$ :

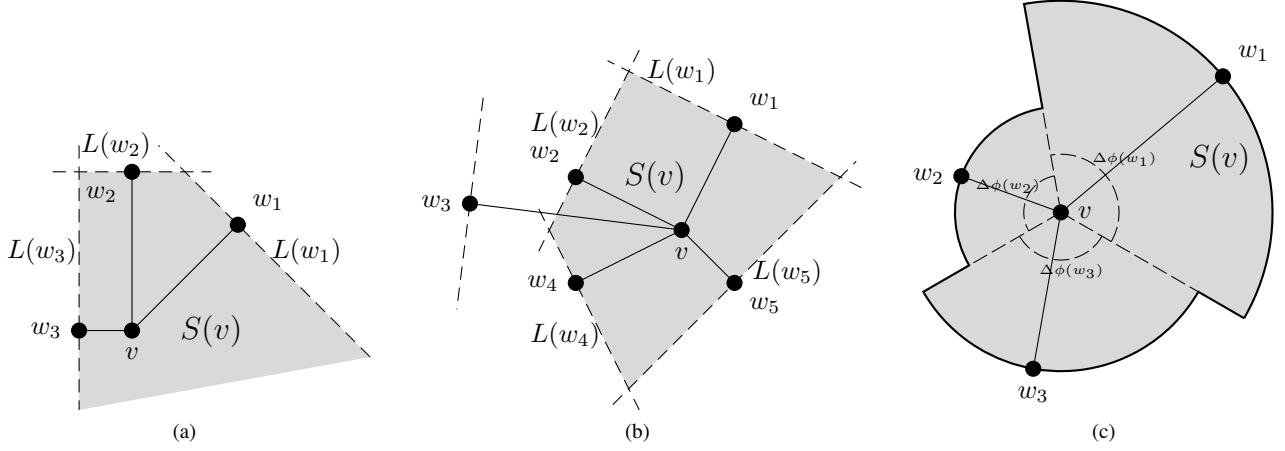


Fig. 4. Deficiencies of original curvature approximation of [12] in (a) and (b) and solution through the new geometric approximation in (c). In (a) the defined region has infinite area, while in (b) neighbor  $w_3$  causes an unintuitive shape for  $S(v)$ . These ill cases are handled properly by defining region  $S(v)$  through the neighbor angles, as done in (c).

$\mathbb{R}^2 \rightarrow \mathbb{R}$  is continuously differentiable and  $\nabla u(v) \neq \mathbf{0}$ , then the approximation of (23) converges in probability to  $\kappa(v)$ .

The full proof of Theorem 3 is given in the Appendix. A brief outline of the proof with its key ideas follows. The main task in the proof is to show that

$$\sum_{i=1}^N I_a(w_i) + I_l(w_i) - \oint_{\Gamma(S(v))} \mathbf{F} \cdot \mathbf{n} \, dl \xrightarrow{P} 0. \quad (24)$$

We treat each component of  $\Gamma(S(v))$ , i.e. each arc and line segment, separately and prove the following convergence in probability results:

$$I_a(w_i) - \int_{C_a(w_i)} \mathbf{F} \cdot \mathbf{n} \, dl \xrightarrow{P} 0$$

and

$$I_l(w_i) - \int_{C_l(w_i)} \mathbf{F} \cdot \mathbf{n} \, dl \xrightarrow{P} 0.$$

Afterwards, these convergence results are combined through the sum property of convergence in probability to obtain (24). To prove the above results, we make use of the continuity of both  $\mathbf{F}$  and  $\nabla u$  at  $v$ , which is ensured by the conditions on  $u$  in Theorem 3. In addition, we use the law of total probability to prove that certain probabilities vanish in the limit, by expanding the examined probability with respect to the mutually exclusive events of  $\mathbf{F}$  being continuous or discontinuous at a point (or on a curve), and exploiting the continuity of  $\mathbf{F}$  at  $v$  to show that the probability of discontinuity vanishes. The last step relies on the fact that  $\rho(n) \in o(1)$ , which implies that the distance between the aforementioned point (or curve) and  $v$  converges in probability to 0.

### B. Gradient-based Approximation

We have seen that the divergence of a vector field can be used to compute the curvature. An alternative way to approximate this divergence is through its differential definition, which avoids handling the geometric quantities of Section IV-A. More specifically, the unit gradient field can

be expressed through its components as  $\mathbf{F} = (F_1, F_2)$ , so that its divergence is

$$\operatorname{div} \mathbf{F} = \frac{\partial F_1}{\partial x} + \frac{\partial F_2}{\partial y}. \quad (25)$$

As a result, a second application of the gradient approximation of (5), this time on the components of the approximate unit gradient field  $\mathbf{F}_g = (F_{1,g}, F_{2,g})$ , is adequate for calculating the curvature. The full expression for this approximation is

$$\kappa(v) \approx \frac{\sum_{i=1}^N \frac{F_{1,g}(w_i) - F_{1,g}(v)}{d(v, w_i)} \cos(\phi(w_i)) \Delta\phi(w_i)}{\pi} + \frac{\sum_{i=1}^N \frac{F_{2,g}(w_i) - F_{2,g}(v)}{d(v, w_i)} \sin(\phi(w_i)) \Delta\phi(w_i)}{\pi}. \quad (26)$$

This gradient-based curvature approximation also converges in probability for RGGs, under slightly stricter conditions than the geometric curvature approximation.

**Theorem 4:** Let  $\mathcal{G}(n, \rho(n))$  be an RGG embedded in  $[0, 1]^2$ , with  $\rho(n) \in \omega(n^{-1/2}) \cap o(1)$  and  $v$  a vertex of  $\mathcal{G}$ . Let  $u : \mathbb{R}^2 \rightarrow \mathbb{R}$  be twice differentiable and  $\nabla u(v) \neq \mathbf{0}$ . Then, the approximation of (26) converges in probability to  $\kappa(v)$ .

*Proof:* With the same argument as in the proof of Theorem 3, it can be shown that

$$\mathbf{F}_g(v) - \mathbf{F}(v) \xrightarrow{P} \mathbf{0}$$

and

$$\mathbf{F}_g(w_i) - \mathbf{F}(w_i) \xrightarrow{P} \mathbf{0} \, \forall i \in \{1, \dots, N\}.$$

These results can be combined into

$$\mathbf{F}_g(w_i) - \mathbf{F}_g(v) - (\mathbf{F}(w_i) - \mathbf{F}(v)) \xrightarrow{P} \mathbf{0} \, \forall i \in \{1, \dots, N\}.$$



Using the above convergence and the product and sum properties of convergence in probability, it follows that

$$\sum_{i=1}^N \frac{F_{1,g}(w_i) - F_{1,g}(v)}{d(v, w_i)} \mathbf{e}_{vw_i} \Delta\phi(w_i) - \sum_{i=1}^N \frac{F_1(w_i) - F_1(v)}{d(v, w_i)} \mathbf{e}_{vw_i} \Delta\phi(w_i) \xrightarrow{P} \mathbf{0}.$$

Furthermore, since  $u$  is twice differentiable and  $\nabla u(v) \neq \mathbf{0}$ ,  $F$  is differentiable at  $v$ , as the quotient of differentiable functions with nonzero denominator. As a result,  $F_1$  is also differentiable at  $v$  and Theorem 1 applies:

$$\frac{\sum_{i=1}^N \frac{F_1(w_i) - F_1(v)}{d(v, w_i)} \mathbf{e}_{vw_i} \Delta\phi(w_i)}{\pi} \xrightarrow{P} \nabla F_1(v).$$

This result can be combined with the previous one through the sum property of convergence in probability to obtain

$$\frac{\sum_{i=1}^N \frac{F_{1,g}(w_i) - F_{1,g}(v)}{d(v, w_i)} \mathbf{e}_{vw_i} \Delta\phi(w_i)}{\pi} \xrightarrow{P} \nabla F_1(v).$$

An identical analysis to the one above leads to the same result for  $F_2$ .

The last step of the proof is to isolate from the last convergence results the  $x$ -component of  $\nabla F_1(v)$  and the  $y$ -component of  $\nabla F_2(v)$ , which appear in (25), and use the sum property of convergence in probability to show that the expression on the right-hand side of (26) converges in probability to  $\partial F_1(v)/\partial x + \partial F_2(v)/\partial y$ . ■

## V. SMOOTHING FILTERING ON GRAPHS

In the last two sections, we developed certain approximations to compute quantities that are essential for the update equation of the GAC model on arbitrary graphs. Despite the convergence of these approximations to the true values of the quantities in the limit of large RGGs, there is a non-negligible error in practice, due to the discrete nature of the approximations. This error is propagated to the embedding function after each update and therefore it may be accumulated after several iterations in a way that leads to instabilities.

Consequently, it is very beneficial to apply smoothing filtering across the graph, in order to increase the robustness of approximations against spatial non-uniformities in vertex locations. Moreover, smoothing filtering is necessary in the initialization stage of the GAC algorithm, where the original intensity function is simplified to distinguish predominant “edges” from small-scale variations. Of course, filtering functions across the graph induces a considerable computational burden, especially when done at every iteration, but we are willing to trade a little speed for stability.

### A. Neighborhood-based Average/Median Filtering

To compute a smoother version of a function defined on a graph, one option is to operate in the same neighborhood-based framework that we presented in the previous sections,

and apply a simple filter on the original version of the function. This filter can be either an average or a median filter, receiving as input the set of function values at the vertex itself and all its neighbors. In the curvature and embedding function cases this is straightforward, while for the gradient, we filter each of the two vector dimensions separately.

To validate the benign effects of smoothing filtering of both the gradient and curvature component of the GAC model, we experiment with certain analytical functions defined on RGGs. For each graph, we compute the function’s gradient and the curvature of its level sets using the proposed approximations and afterwards we filter the results with an average or median filter. Deriving the analytical expressions of the function’s gradient and curvature, we are able to compare them with our estimates. To enable a quantitative assessment of our approximations and smoothing filters, we define a suitable error metric which we call *relative error* and denote by  $e_r$ . If the approximation error at each vertex is defined as the difference between the approximate value of the function at that vertex and its true, analytical value, then the relative error on the whole graph is simply the ratio of the error’s energy and the function’s energy:

$$e_r = \frac{E_{\text{error}}}{E_{\text{analytical}}}. \quad (27)$$

In the experiments that follow in the rest of the paper, the radius of an RGG is computed as

$$\rho(n) = Cn^{-1/3}, \quad (28)$$

in order to achieve the strictest asymptotic bound for gradient approximation error according to the results of Section III. Unless otherwise specified, we fix  $C = 0.6$  in the rest of the paper. In the rest of this section, all RGGs are embedded in  $[0, 1]^2$ . In Fig. 6, we show the results of gradient approximation for an isotropic Gaussian on an RGG with  $n = 5500$  vertices. The analytical form of the Gaussian is

$$\exp\left(-((x - x_0)^2 + (y - y_0)^2) / 2\sigma^2\right), \quad (29)$$

with  $\sigma = 0.25$  and  $x_0 = y_0 = 0.5$ . The small, local deviations of the geometric approximation (Fig. 6(b)) from the true gradient vector field (Fig. 6(a)) are almost completely smoothed out with average or median filtering (Fig. 6(c) and 6(d) respectively).

To verify the enhancement of gradient approximation with smoothing filtering quantitatively, we evaluate the relative error for the Gaussian function in (29) on RGGs whose size ranges from 1000 to 10000 vertices. Fig. 7 shows average values of  $e_r$  over 10 different graphs for each size, which leads to a reduced variance in the estimation. Using either smoothing filter reduces relative error substantially irrespective of size. This leads us to apply smoothing on gradient and feed the smoothed version to curvature computation.

Fig. 8 presents a comparison of different curvature approximations for a function whose graph corresponds to an elliptical cone, which we will call conic function for short. The form of this conic function is

$$\sqrt{\frac{(x - x_0)^2}{\alpha^2} + \frac{(y - y_0)^2}{\beta^2}}, \quad (30)$$

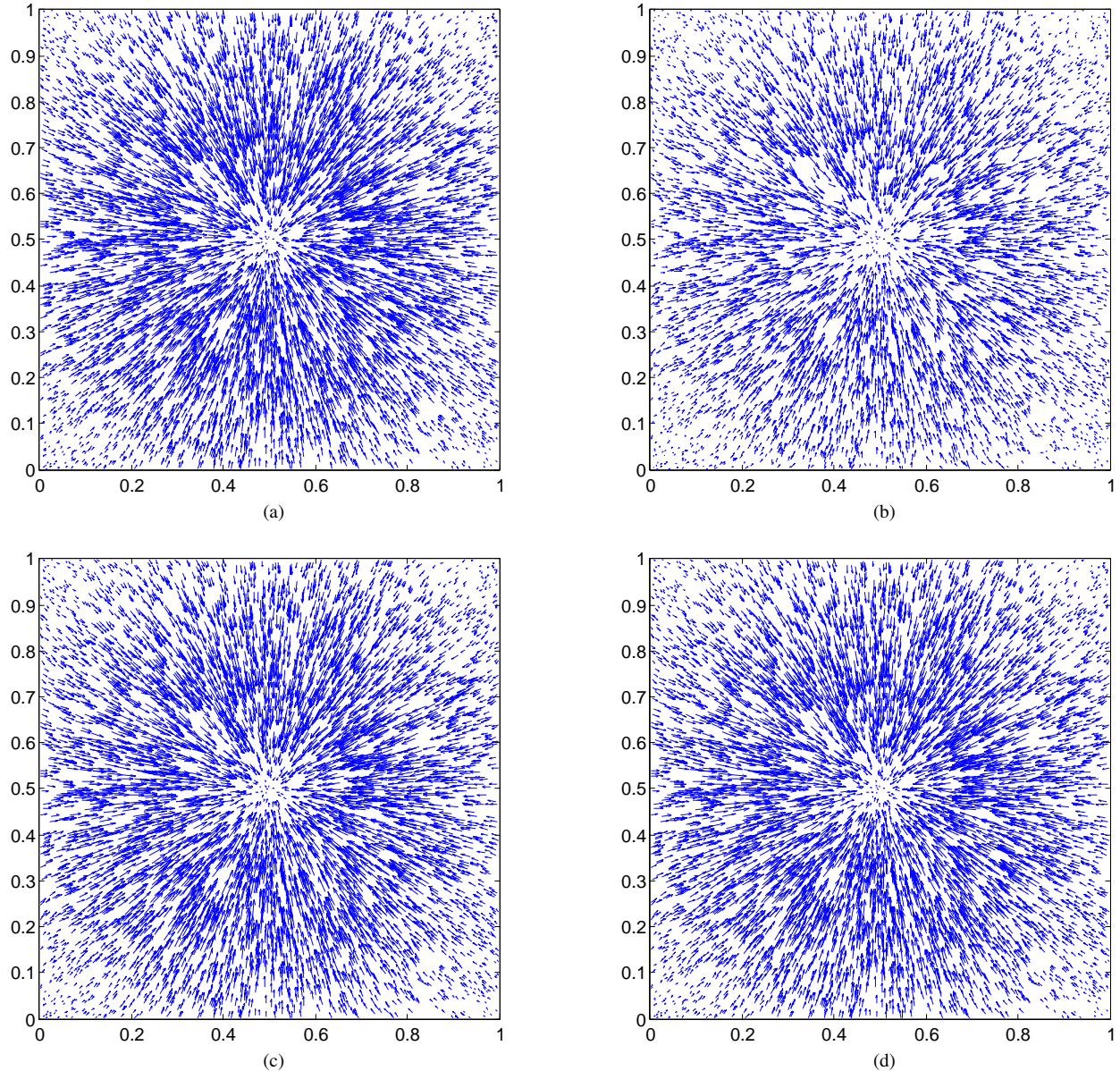


Fig. 6. Comparison of gradient approximations for a Gaussian function defined on an RGG. The true gradient field is shown in (a). The result of the geometric approximation is shown in (b). In (c) and (d), average and median filtering have been used respectively to improve (b). All vector fields have been scaled by a factor of 2 to aid visualization.

where  $\alpha = 0.4$ ,  $\beta = 0.3$ ,  $x_0 = -0.25$  and  $y_0 = 0.5$  for Fig. 8. The underlying graph is an RGG with 7000 vertices. Even though all approximations are smoothed, they demonstrate strong, abrupt variations from the true curvature values. This degradation relative to the gradient case is explained by the cascaded nature of curvature approximation. Nonetheless, the overall curvature trend of the conic function is captured well by all approaches.

In Fig. 9(a), we present the results of an experiment similar to the one in Fig. 7, this time focusing on the curvature of the conic function in (30). The parameters of the function are the same as before. Median filtering appears superior: the median-filtered approximations exhibit lower error than the corresponding average-filtered ones over almost the entire range of graph sizes (except for the smaller sizes). Even more

importantly, the relative error of median-filtered curvature is steadily decreasing for increasing graph size both with the geometric and the gradient-based approximation, in contrast to the average-filtered cases, where the error stops decreasing around 4000 vertices. Due to these facts, we use median filtering for smoothing gradient and curvature for the GAC model in the rest of the paper. We also observe that the geometric approximation induces a relatively smaller error than the gradient-based approximation, when the same filter is used.

To emphasize on the importance of the assumptions made for convergence of the curvature approximations in Section IV, we repeat the last experiment setting  $x_0 = 0.5$ . This way, the conic function is not differentiable at  $(0.5, 0.5)$ , which lies in the interior of the graphs' region, and therefore the

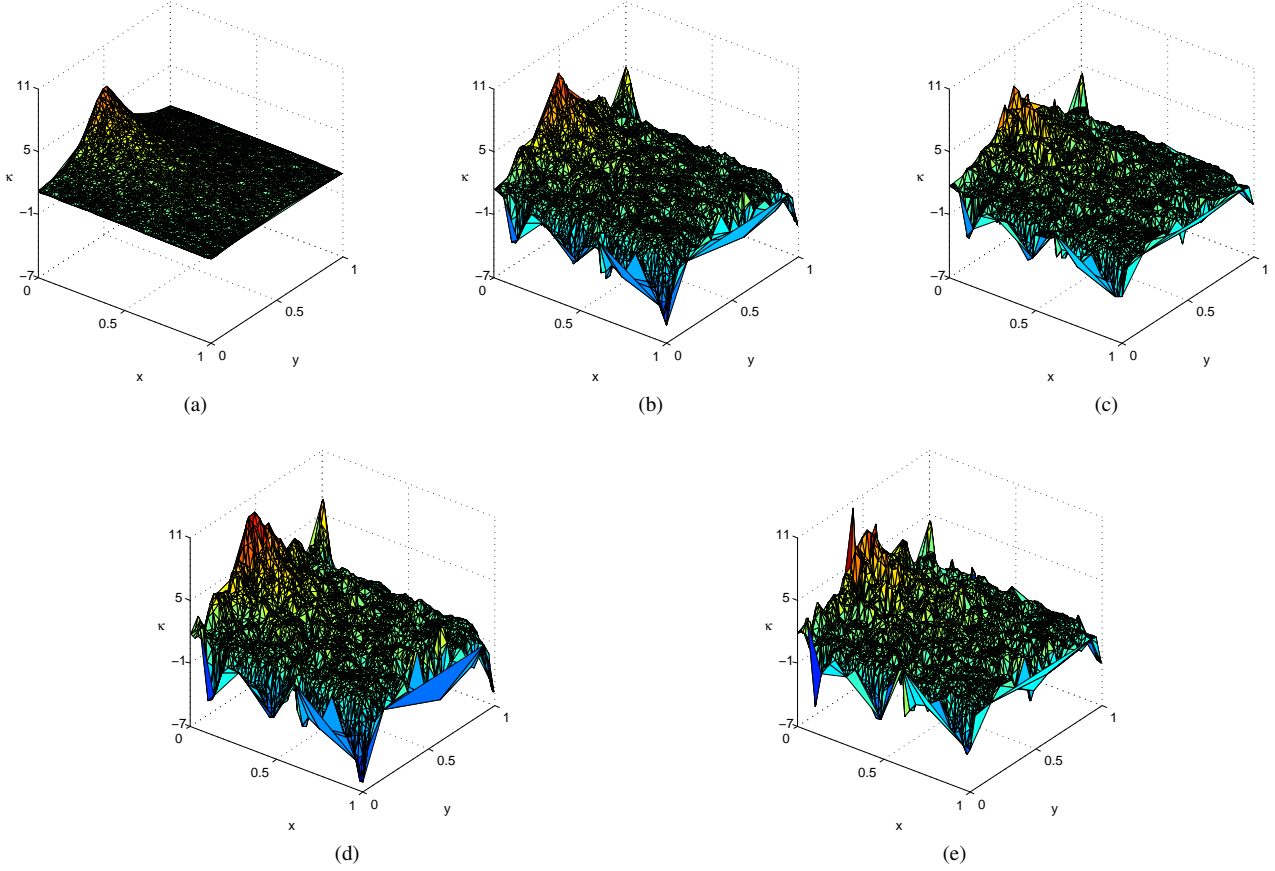


Fig. 8. Comparison of curvature approximations for a conic function defined on an RGG. The true curvature is shown in (a). (b) and (c) show the geometric approximation smoothed with average and median filtering respectively, while (d) and (e) show the gradient-based approximation smoothed with average and median filtering respectively. In each case, the smoothing filter used for curvature has also been applied to the input gradient values.

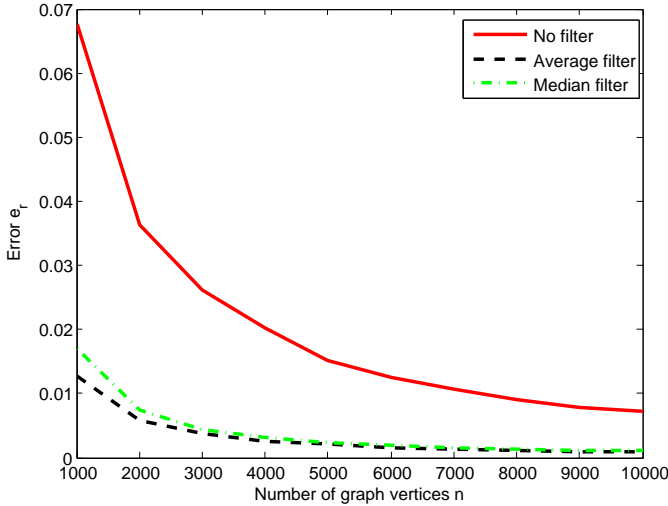


Fig. 7. Relative error of gradient approximations for a Gaussian function defined on RGGs of increasing size. The geometric approximation with no filtering is compared to its filtered versions with an average or median filter.

assumptions of Theorems 3 and 4 do not hold necessarily for every vertex of the graphs. In fact, near this point, the true curvature of the level sets approaches infinity. Indeed, the evolution of relative error depicted in Fig. 9(b) confirms that

all approximations are less accurate and they do not converge in this case.

### B. Gaussian Smoothing

At the initialization stage of the GAC algorithm, one of the tasks is to process the original image function  $I : \mathcal{V} \rightarrow \mathbb{R}^+$  so as to obtain a smoother version of it. This way, the stopping function  $g$  can be computed subsequently, encoding only the predominant edges and ignoring local variations on the graph.

Following [12], we employ Gaussian smoothing defined on graphs for this task. The filter is an isotropic 2D Gaussian with standard deviation  $\sigma$ :

$$G_\sigma(\mathbf{x}) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{\|\mathbf{x}\|^2}{2\sigma^2}\right). \quad (31)$$

We will denote the smoothed image which is obtained by using such a filter with  $I_\sigma$ . The authors in [12] use a simple graph-based convolution of (31) with the image to perform smoothing:

$$I_\sigma(\mathbf{v}) = \sum_{\mathbf{w} \in \mathcal{V}} I(\mathbf{w}) G_\sigma(\mathbf{v} - \mathbf{w}). \quad (32)$$

A simple formula is then used to calculate the stopping function  $g$ :

$$g(\|\nabla I_\sigma\|) = \frac{1}{1 + \frac{\|\nabla I_\sigma\|^2}{\lambda^2}}. \quad (33)$$

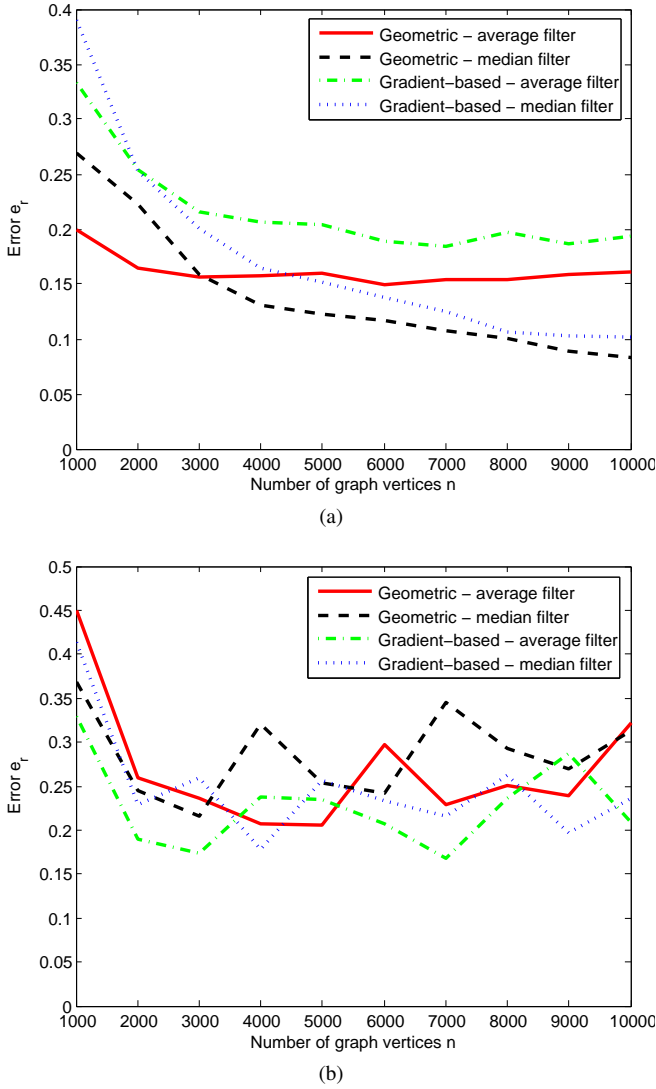


Fig. 9. Relative error of curvature approximations for a conic function defined on RGGs of increasing size. In (a), the conic function does not assume a local extremum in the interior of the graph’s region, whereas in (b) it does. All four combinations of type of approximation (geometric or gradient-based) and smoothing filter (average or median) are compared in both cases.

However, the arbitrary graph setting introduces non-uniformities: in some parts of the graph, the vertices might be distributed more densely than in other parts. This implies that (32) will operate counter-intuitively, introducing variations to the smoothed image in regions of the graph where the original intensity function is constant. To demonstrate this behavior, we use a simple binary image of a disk, shown in Fig. 10(a). The result of plain Gaussian filtering of this image is shown in Fig. 10(b). Not only has the range of image values changed, but also the interior of the original disk now exhibits significant variations in intensity values. This shortcoming is propagated to  $\|\nabla I_\sigma\|$  and  $g$  values, as we present in Fig. 10(e) and Fig. 10(h) respectively. There is a deviation of  $g$  values from the ideal value of 1 inside the area corresponding to the disk and a variation in these values as well, which means that the gradient of the stopping function is not  $\mathbf{0}$ , as it should.

To tackle this issue, we add a normalization term to (32) to

account for non-uniformities:

$$I_\sigma(\mathbf{v}) = \frac{\sum_{\mathbf{w} \in \mathcal{V}} I(\mathbf{w}) G_\sigma(\mathbf{v} - \mathbf{w})}{\sum_{\mathbf{w} \in \mathcal{V}} G_\sigma(\mathbf{v} - \mathbf{w})}. \quad (34)$$

We term this method normalized Gaussian filtering and show its result for the examined disk image in Fig. 10(c). The smoothed image is now very similar to the output of simple Gaussian filtering in the usual image processing setting with regularly spaced pixels. As a result, the corresponding magnitude of the gradient of  $I_\sigma$  and  $g$  function (shown in Fig. 10(f) and Fig. 10(i) respectively) match our expectations.

An important observation at this point is that in the stopping function computation pipeline, we are rather interested in the smoothed image’s gradient than in the smoothed image itself. Since the derivatives of the Gaussian filter have closed analytical forms, it is appealing to exchange the convolution with the gradient operator and convolve the image directly with Gaussian derivatives in order to obtain the gradient of  $I_\sigma$ . In this case, normalization is not straightforward as in normalized Gaussian filtering: Gaussian derivatives assume both positive and negative values. We circumvent this difficulty by splitting the vertices into two sets, according to the sign of the Gaussian derivative with respect to the processed vertex, and perform separate normalization for each of these sets. This separation can be easily expressed in terms of the vertices’ coordinates. If we denote  $\mathbf{v} = (v_1, v_2)$ , then Gaussian derivative filtering with separate normalization is defined as:

$$\nabla I_\sigma(\mathbf{v}) = \left[ \begin{array}{c} \frac{\sum_{\substack{\mathbf{w} \in \mathcal{V}: \\ w_1 \geq v_1}} I(\mathbf{w}) \frac{\partial G_\sigma(\mathbf{v} - \mathbf{w})}{\partial x}}{\sum_{\substack{\mathbf{w} \in \mathcal{V}: \\ w_1 \geq v_1}} \frac{\partial G_\sigma(\mathbf{v} - \mathbf{w})}{\partial x}} + \frac{\sum_{\substack{\mathbf{w} \in \mathcal{V}: \\ w_1 < v_1}} I(\mathbf{w}) \frac{\partial G_\sigma(\mathbf{v} - \mathbf{w})}{\partial x}}{-\sum_{\substack{\mathbf{w} \in \mathcal{V}: \\ w_1 < v_1}} \frac{\partial G_\sigma(\mathbf{v} - \mathbf{w})}{\partial x}} \\ \frac{\sum_{\substack{\mathbf{w} \in \mathcal{V}: \\ w_2 \geq v_2}} I(\mathbf{w}) \frac{\partial G_\sigma(\mathbf{v} - \mathbf{w})}{\partial y}}{\sum_{\substack{\mathbf{w} \in \mathcal{V}: \\ w_2 \geq v_2}} \frac{\partial G_\sigma(\mathbf{v} - \mathbf{w})}{\partial y}} + \frac{\sum_{\substack{\mathbf{w} \in \mathcal{V}: \\ w_2 < v_2}} I(\mathbf{w}) \frac{\partial G_\sigma(\mathbf{v} - \mathbf{w})}{\partial y}}{-\sum_{\substack{\mathbf{w} \in \mathcal{V}: \\ w_2 < v_2}} \frac{\partial G_\sigma(\mathbf{v} - \mathbf{w})}{\partial y}} \end{array} \right] \quad (35)$$

The application of Gaussian derivative filtering with separate normalization on the examined image produces the results shown in Fig. 10(d) for  $\|\nabla I_\sigma\|$  and Fig. 10(g) for  $g$ . The quality of the stopping function is at least as satisfactory as in the normalized Gaussian filtering case of Fig. 10(i). Consequently, both our novel methods for Gaussian smoothing on graphs outperform the simple Gaussian filtering approach and can be readily used in the GAC framework.

## VI. RESULTS

Having approximated the various terms of the active contour model in the arbitrary graph setting, we are able to apply the iterative algorithm for object detection on graphs that stems



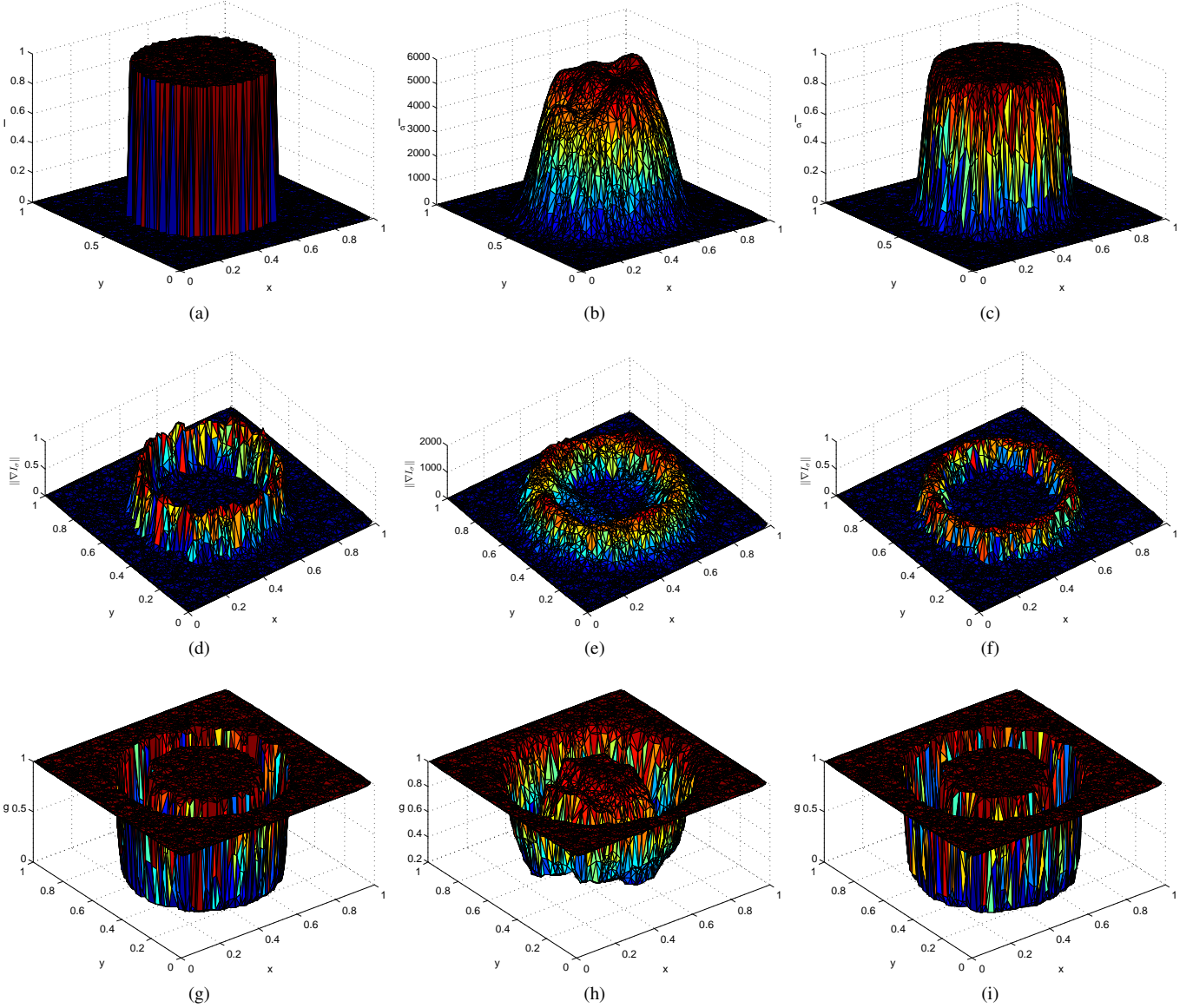


Fig. 10. Comparison of methods for Gaussian smoothing and computation of stopping function. The original image on the graph (a disk) is shown in (a). The rest of the figure is organized as follows: the two rightmost plots of the top row contain smoothed versions  $I_\sigma$  of the original image, the middle row contains gradient magnitudes of  $I_\sigma$  and the bottom row contains  $g$  values based on these gradient magnitudes. The results in the left column pertain to Gaussian derivative filtering with separate normalization using  $\sigma = 0.02$  and  $\lambda = 0.05$ , those in the middle column pertain to simple Gaussian filtering using  $\sigma = 0.05$  and  $\lambda = 1000$  and those in the right column correspond to normalized Gaussian filtering with  $\sigma = 0.02$  and  $\lambda = 0.05$ . We use the approximation of (17) to compute the gradient magnitude in (e) and (f). For all three methods, we filter  $\|\nabla I_\sigma\|$  with a median filter before feeding it to the formula for  $g$ .

from the relevant PDE. The input comprises a graph  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  whose vertices are embedded in  $\mathbb{R}^2$  and a real-valued function  $I : \mathcal{V} \rightarrow \mathbb{R}^+$ . The algorithm includes the following steps:

- 1) Compute  $g(\|\nabla I_\sigma\|)$ , using either (34) and (17) or (35) to compute  $\|\nabla I_\sigma\|$ . In both cases, median filtering is applied to  $\|\nabla I_\sigma\|$  before plugging it into the formula for  $g$ . Then, compute the magnitude of  $g$ 's gradient using (17) and its direction with (5).
- 2) Choose a subset  $X$  of  $\mathcal{V}$  which contains the objects to be detected and initialize the embedding function with the signed distance function from the boundary of  $X$ , denoted by  $u_0$ . By convention,  $u_0$  is positive inside  $X$ .

3) Iterate for  $r \in \mathbb{N}$

$$u_r = u_{r-1} + \Delta t ((\kappa - c) \|\nabla u_{r-1}\| g + \nabla g \cdot \nabla u_{r-1}) \quad (36)$$

until convergence, i.e. until  $u$  has not changed its sign at most vertices for several consecutive iterations. In the difference equation (36),  $\Delta t$  and  $c$  are positive constants and  $\kappa$  is the curvature of the level sets of  $u_{r-1}$ .

In practice, after each iteration of step 3 of the algorithm, we smooth  $u_r$  with a median filter before proceeding to the next iteration. The parameters involved in the algorithm are the time step  $\Delta t$  of the difference equation, the balloon force constant  $c$ , the scale  $\sigma$  of the Gaussian smoothing filter and parameter  $\lambda$  in  $g$ 's formula. Tuning their values depending on the particular input is pivotal in obtaining satisfactory segmentation results.

In the following experiments, unless otherwise specified, we set  $\Delta t = 0.005$ ,  $c = 20$ ,  $\sigma = 0.02$  and  $\lambda = 0.05$ .

Another important aspect in applying the active contour model on graphs is the method used to create the graph. The original input often consists only of a set of intensity values at vertex locations, without any information about the edges of the graph. This setting leaves us free to choose the underlying model for the structure of the graph. In our experiments, we used random geometric graphs and Delaunay triangulations (DTs). In other cases, one may have a full image at her disposal; however, the graph framework is still relevant. In particular, sampling the image uniformly at random with much fewer samples than the total number of pixels brings us to the previous setting and at the same time reduces the size of the input compared to the standard image-based active contour framework. An attractive alternative to random sampling is to extract vertex locations via watershed transformation. More specifically, we apply watershed transformation directly to the gradient of the image and place the vertices at the centroids (ultimate erosions) of the resulting segments. This approach leads to far better detection results than randomly sampling the image, as it captures image particularities into the structure of the graph and “compresses” the intensity function to the part that is crucial for segmentation.

First, we apply our method to synthetic data. In particular, we construct RGGs by placing  $n$  vertices uniformly at random in  $[0, 1]^2$  and using (28) to create the edges. We then define binary image functions on the graphs, which model simple shapes. Fig. 11 shows the evolution of the active contour when detecting a triangle, with  $n = 5500$ . The final form of the contour in Fig. 11(g), comprising the green and red vertices, does not capture well the sharp corners of the triangle, which is expected due to the isotropic Gaussian smoothing of the image.

A harder benchmark is a non-connected shape, such as the two disks of Fig. 12, where  $n = 5500$ . In order to allow the contour to change its topology and separate the disks, we need to tune the spatial parameters of the algorithm based on the distance of the objects. Specifically, the smoothed objects’ boundaries must be at least two radii apart for  $g$  to assume values close to 1 between the objects, which is a necessary condition for separation. Setting  $\sigma = 0.005$ ,  $\lambda = 0.1$ ,  $C = 0.45$  and  $c = 40$ , this condition is satisfied (Fig. 12(b)) and the contour is able to distinguish the two disks (Fig. 12(f)).

Our next experiment involves a full grayscale image with four distinct coins (Fig. 13(a)). We make a two-fold comparison, on the one hand between placing the graph’s vertices at random or via watershed transformation and on the other hand between using random geometric or DT structure for the edges. Results from the four experiments corresponding to all possible combinations are shown in Fig. 13(c)–(f). To ensure a fair comparison, the number of randomly placed vertices is approximately the same as in the watershed case. In all the experiments, we set  $c = 2$ . The most accurate segmentation is achieved with DT and watershed-placed vertices (Fig. 13(f)), as the boundaries of the objects are captured very well. Using randomly placed vertices and DT structure, or watershed and random geometric structure, also yields decent results (Fig.

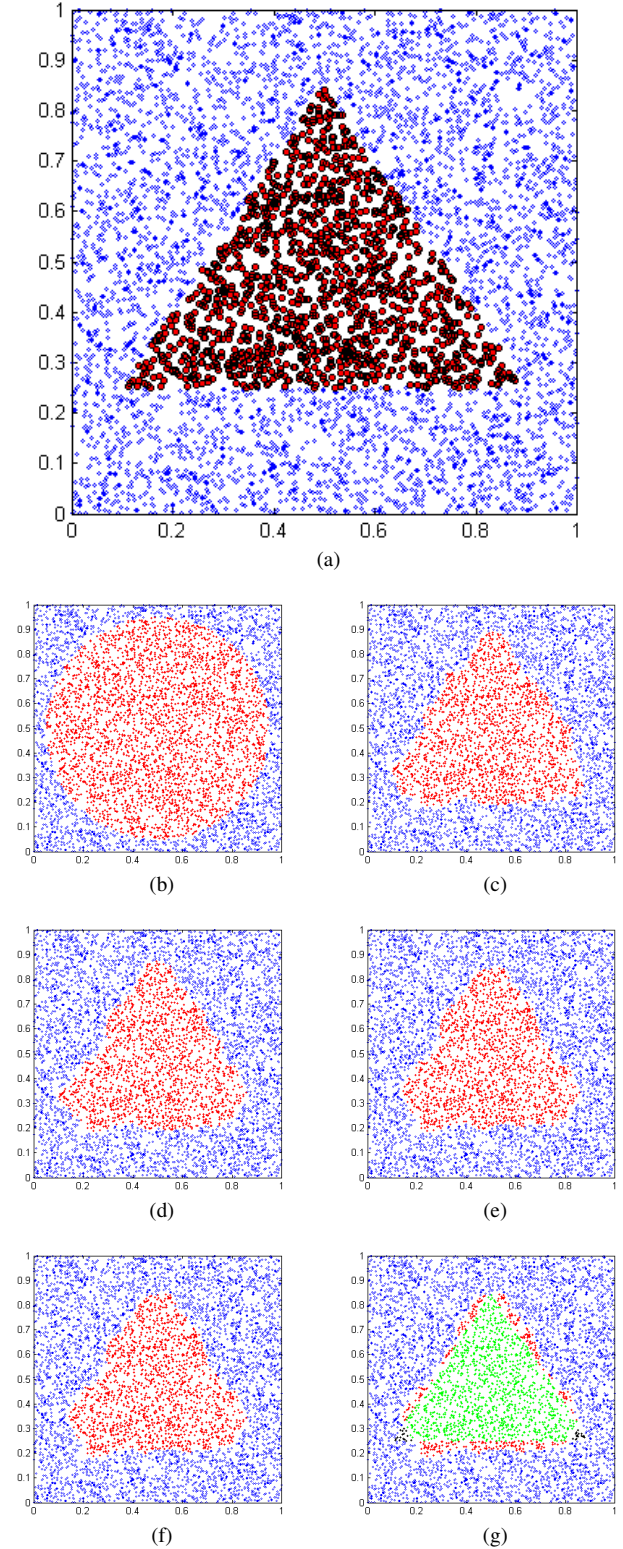


Fig. 11. Detection of triangle on a random geometric graph. Edges are omitted for illustration purposes. (a) Original triangle on graph (b)–(f) Instances of active contour evolution at intervals of 60 iterations, with vertices in the contour’s interior shown in red and the rest in blue (g) Final detection result after 300 iterations, using green for true positives, blue for true negatives, red for false positives and black for false negatives.



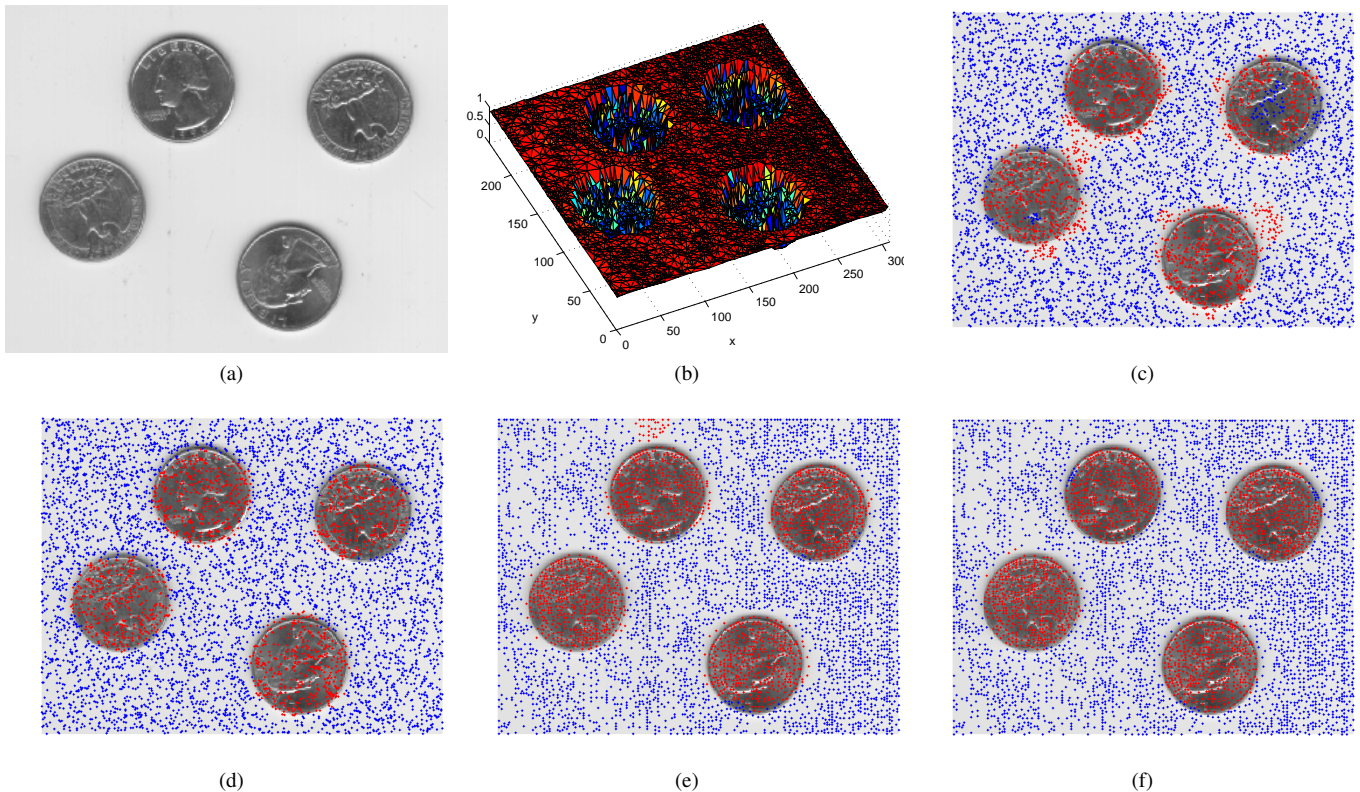


Fig. 13. Detection of grayscale objects on graphs with active contours. (a) Full grayscale image with four coins (b) Intensity function on watershed-based DT (c)–(f) Final detection results overlaid on original image, with detected objects shown in red and background in blue. We use (c) randomly placed vertices with random geometric structure,  $\sigma = 0.02$  and  $\lambda = 0.03$ , (d) randomly placed vertices with DT structure,  $\sigma = 0.02$  and  $\lambda = 0.03$ , (e) watershed-placed vertices with random geometric structure,  $\sigma = 0.008$  and  $\lambda = 0.07$ , and (f) watershed-placed vertices with DT structure,  $\sigma = 0.01$  and  $\lambda = 0.07$ . The total number of iterations to obtain the final segmentation result is (c) 2200, (d) 4000, (e) 3000 and (f) 4000.

13(d) and (e)). On the contrary, the combination of randomly placed vertices and random geometric structure (i.e. a proper RGG) leads to poor segmentation, in which close objects are not separated and others have holes opened in their interior (Fig. 13(c)). Consequently, the DT structure is preferable to the random geometric one and usage of watershed transformation to place vertices when a full image is available is better than random placement.

We use the combination of watershed-placed vertices and DT structure for the edges to repeat the above experiment for a collection of natural color images coming from the Berkeley Segmentation Dataset BSDS500 [19]. The images were converted to grayscale for the application of our method. Segmentation results on the images are presented in Fig. 14. In general, our algorithm detects the dominant objects in the images successfully, even though background clutter and thin protrusions or concavities of the objects' boundaries may cause minor inaccuracies.

An interesting application of our method is related to geographical data, where the two spatial coordinates are longitude and latitude and the image function can encode information about any type of real-valued signal defined at the vertices of the graph. Such a signal is the average annual wind speed, which is particularly important for locating regions with high wind power potential. Fig. 15 presents the application of our active contour algorithm to average annual wind speed data

on a graph constructed as a Delaunay triangulation. We use  $\sigma = 0.05$ ,  $\lambda = 0.8$  and  $c = 5$  and normalize the coordinates and the values of wind speed. The algorithm detects a cluster that corresponds to a region with relatively uniform and quite high wind speed. Another example regards the signal strength of a cellular network. Fig. 16 demonstrates the result of our method for such data on a Delaunay triangulation. We again normalize the data and use a very small time step  $\Delta t = 10^{-4}$  to guarantee convergence, which requires far more iterations until termination than in the previous experiments. In addition, we set  $\sigma = 0.03$  and  $\lambda = 0.02$ . The segmented set of vertices comprises two regions, the southern of which is characterized by an increased signal strength compared to the rest of the graph. Our approach is tailored for geographical data with arbitrary spatial configuration such as the above cases, which grants greater flexibility when collecting measurements.

## VII. CONCLUSION

In this paper, we introduce approximations of the gradient and curvature terms involved in the level set formulation of active contour evolution models for the case of arbitrary graphs. We examine theoretically the conditions under which these approximations converge in probability to the true value of the operators and, in the case of gradient, the respective rate of convergence, in the limit of large graphs. At an algorithmic level, we propose smoothing filtering to improve accuracy of

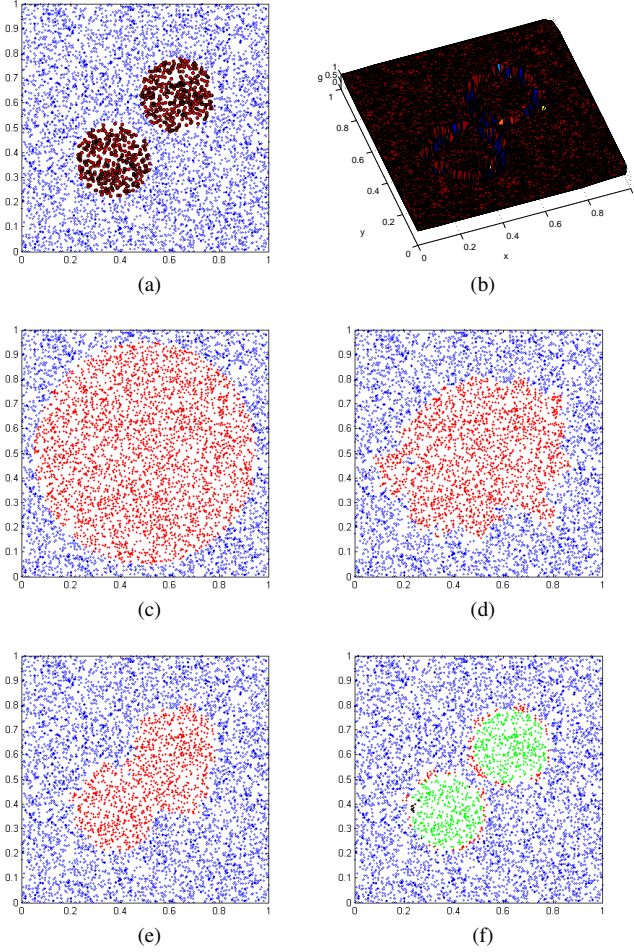


Fig. 12. Detection of two disks on a random geometric graph. (a) Original disks on graph (b)  $g$  function (c)–(e) Instances of active contour evolution at intervals of 40 iterations, with vertices in the contour's interior shown in red and the rest in blue (f) Final detection result after 120 iterations, using green for true positives, blue for true negatives, red for false positives and black for false negatives.

such approximations and help the partial difference scheme (36) converge properly, and provide improved implementations of Gaussian smoothing on graphs which account for potential non-uniformities. We also demonstrate the applicability of active contours on graphs, equipped with our approximations, for segmentation of regular images as well as raw geographical data.

A remaining challenge in our work is related to the curvature term of active contour models. Both the geometric and the gradient-based approximation which are proposed are proved to converge in probability to the true value of curvature at points with nonzero gradient, which is the first result of this kind to the best of our knowledge. However, due to the cascaded approximation that we perform, a larger amount of noise is injected in the approximate values, which is also reflected in the slower convergence observed in our empirical tests compared to gradient approximation. This forces us to take a very small step in time in some cases when updating the embedding function, which leads to much slower convergence of the active contour algorithm. To overcome this difficulty, a deeper analysis of the curvature term needs to be

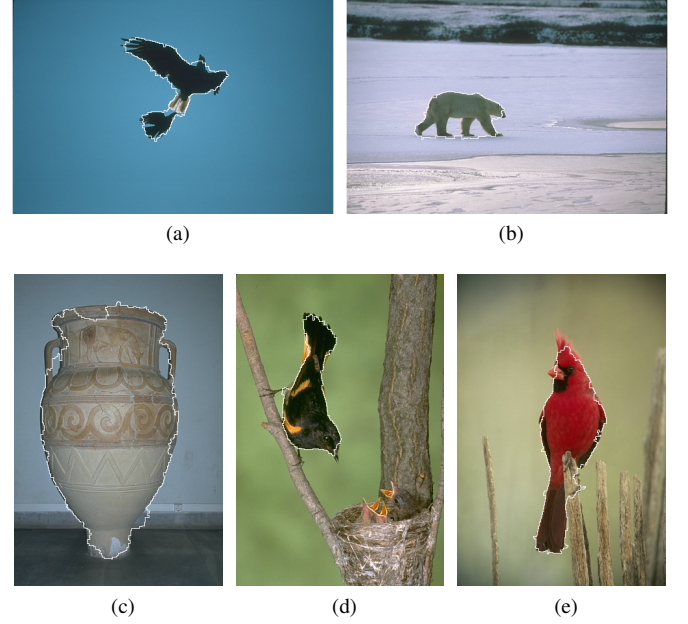


Fig. 14. Segmentation of images from BSDS500 dataset. The detected boundary, marked white, is defined as the boundary of the union of the watershed segments for which the corresponding graph vertex belongs to the final contour's interior. We use  $\sigma = 0.005$ ,  $\lambda = 0.03$  and  $c = 2$  in all cases except (d), for which we set  $\lambda = 0.05$ , and (e), for which we set  $\sigma = 0.01$  and  $\lambda = 0.02$ . The total number of iterations to obtain the final segmentation result is (a) 3400, (b) 2000, (c) 2600, (d) 4400 and (e) 5600.

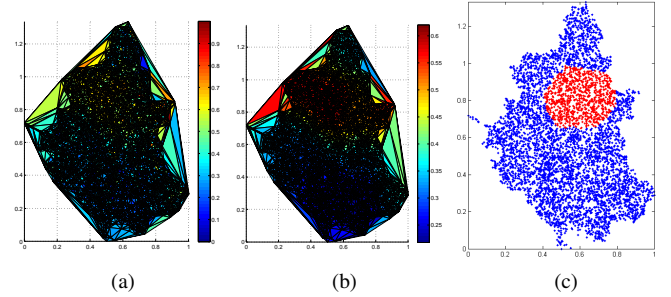


Fig. 15. Segmentation of wind speed data on a graph. (a) Normalized data on graph (b) Smoothed wind speed (c) Final detection result after 1000 iterations, with vertices in the contour's interior shown in red and the rest in blue.

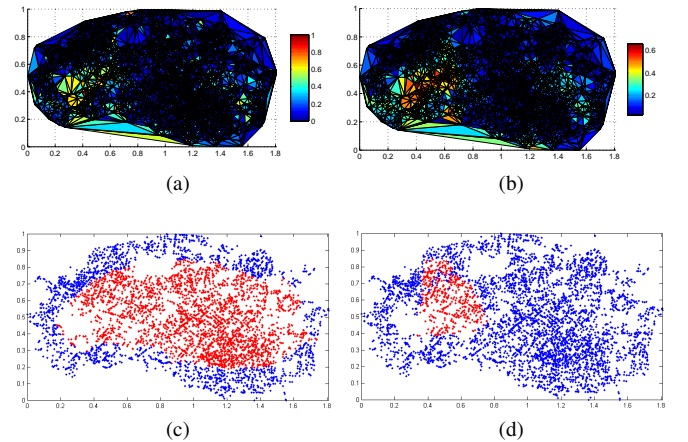


Fig. 16. Segmentation of signal strength data of a cellular network on a graph. (a) Normalized data on graph (b) Smoothed signal strength (c) Initial contour (d) Final detection result after 40000 iterations, with vertices in the contour's interior shown in red and the rest in blue.



accomplished, ideally establishing asymptotic bounds on the respective approximation error similar to our bound for the gradient approximation error.

Our theoretical analysis considers random geometric graphs, whose definition simplifies convergence proofs for our approximations. However, judging from segmentation quality, more regular graph structures, such as Delaunay triangulations, lead to more accurate results. Therefore, an interesting extension of our work is the theoretical study of our approximations on graphs formed as Delaunay triangulations. Furthermore, our initial Gaussian smoothing of the image corresponds to isotropic diffusion, which blurs predominant edges and rounds corners, like in Fig. 11. More faithful preservation of edges in the final segmentation can be ensured by defining anisotropic smoothing on graphs.

#### APPENDIX

*Proof of Theorem 3:* Due to differentiability of  $u$ , Theorem 1 implies that the difference between the geometric gradient approximation  $\nabla u_g(w)$  of (5) and the true value of the gradient  $\nabla u(w)$  converges in probability to zero for every neighbor  $w$  of  $v$ :

$$\nabla u_g(w) - \nabla u(w) \xrightarrow{P} \mathbf{0}.$$

We show that a consequence of this convergence is that  $\mathbf{F}_g(w) - \mathbf{F}(w) \xrightarrow{P} \mathbf{0}$ . The law of total probability implies that for every  $\epsilon > 0$

$$\begin{aligned} & \mathbb{P}(\|\mathbf{F}_g(w) - \mathbf{F}(w)\| > \epsilon) \\ & \leq \mathbb{P}(\nabla u(w) = \mathbf{0}) + \mathbb{P}(\nabla u_g(w) = \mathbf{0} \mid \nabla u(w) \neq \mathbf{0}) \\ & + \mathbb{P}(\|\mathbf{F}_g(w) - \mathbf{F}(w)\| > \epsilon \mid \nabla u(w) \neq \mathbf{0} \text{ and } \nabla u_g(w) \neq \mathbf{0}). \end{aligned}$$

We examine the terms on the right-hand side of the above inequality, showing that all three of them converge to 0, which guarantees that their sum converges to 0 as well. Since  $\|\nabla u(v)\| > 0$ , it holds for the first term that

$$\begin{aligned} \mathbb{P}(\nabla u(w) = \mathbf{0}) & \leq \mathbb{P}\left(\|\nabla u(w)\| < \frac{\|\nabla u(v)\|}{2}\right) \\ & = \mathbb{P}\left(\|\nabla u(v)\| - \|\nabla u(w)\| > \frac{\|\nabla u(v)\|}{2}\right) \\ & \leq \mathbb{P}\left(\|\nabla u(w) - \nabla u(v)\| > \frac{\|\nabla u(v)\|}{2}\right) \rightarrow 0, \end{aligned}$$

where the last limit is due to  $\nabla u(w) \xrightarrow{P} \nabla u(v)$ , as  $\mathbf{w} \xrightarrow{P} \mathbf{v}$  and  $\nabla u$  is continuous. For the second term, the sum property of convergence in probability yields

$$\nabla u_g(w) - \nabla u(w) + \nabla u(w) = \nabla u_g(w) \xrightarrow{P} \nabla u(v)$$

and hence this term can be bounded as follows:

$$\begin{aligned} \mathbb{P}(\nabla u_g(w) = \mathbf{0}) & \leq \mathbb{P}\left(\|\nabla u_g(w)\| < \frac{\|\nabla u(v)\|}{2}\right) \\ & = \mathbb{P}\left(\|\nabla u(v)\| - \|\nabla u_g(w)\| > \frac{\|\nabla u(v)\|}{2}\right) \\ & \leq \mathbb{P}\left(\|\nabla u_g(w) - \nabla u(v)\| > \frac{\|\nabla u(v)\|}{2}\right) \rightarrow 0. \end{aligned}$$

In the third term, it is given that  $\|\nabla u_g(w)\| \geq \delta > 0$  for some  $\delta$ . We write

$$\begin{aligned} \mathbf{F}_g(w) - \mathbf{F}(w) & = \frac{\nabla u_g(w) \|\nabla u(w)\| - \nabla u(w) \|\nabla u_g(w)\|}{\|\nabla u_g(w)\| \|\nabla u(w)\|} \\ & = (\|\nabla u(w)\| (\nabla u_g(w) - \nabla u(w)) \\ & + \nabla u(w) (\|\nabla u(w)\| - \|\nabla u_g(w)\|)) / (\|\nabla u_g(w)\| \|\nabla u(w)\|). \end{aligned}$$

The triangle inequality can be applied to the above expression:

$$\begin{aligned} \|\mathbf{F}_g(w) - \mathbf{F}(w)\| & \leq \left( \|\nabla u(w)\| \|\nabla u_g(w) - \nabla u(w)\| \right. \\ & + \|\nabla u(w)\| \left| \|\nabla u(w)\| - \|\nabla u_g(w)\| \right| \Big) \\ & / (\|\nabla u_g(w)\| \|\nabla u(w)\|) \\ & \leq \frac{2 \|\nabla u_g(w) - \nabla u(w)\|}{\|\nabla u_g(w)\|} \leq \frac{2 \|\nabla u_g(w) - \nabla u(w)\|}{\delta}. \end{aligned}$$

Therefore, given that  $\nabla u(w) \neq \mathbf{0}$  and  $\nabla u_g(w) \neq \mathbf{0}$ , it holds for every  $\epsilon > 0$  that

$$\begin{aligned} & \mathbb{P}(\|\mathbf{F}_g(w) - \mathbf{F}(w)\| > \epsilon) \\ & \leq \mathbb{P}\left(\frac{2 \|\nabla u_g(w) - \nabla u(w)\|}{\delta} > \epsilon\right) \\ & = \mathbb{P}\left(\|\nabla u_g(w) - \nabla u(w)\| > \frac{\epsilon \delta}{2}\right) \rightarrow 0. \end{aligned}$$

As a result,

$$\mathbf{F}_g(w) - \mathbf{F}(w) \xrightarrow{P} \mathbf{0}. \quad (37)$$

Using the closed form (21) for  $I_a(w_i)$  and the product and sum properties of convergence in probability, we obtain

$$\int_{C_a(w_i)} (\mathbf{F}_g(w_i) - \mathbf{F}(w_i)) \cdot \mathbf{n} \, d\ell \xrightarrow{P} 0. \quad (38)$$

In addition, the law of total probability implies that for every  $\epsilon > 0$

$$\begin{aligned} & \mathbb{P}\left(\left|\int_{C_a(w_i)} (\mathbf{F}(w_i) - \mathbf{F}) \cdot \mathbf{n} \, d\ell\right| > \epsilon\right) \\ & \leq \mathbb{P}(\mathbf{F} \text{ not continuous at some point on } C_a(w_i)) \\ & + \mathbb{P}\left(\left|\int_{C_a(w_i)} (\mathbf{F}(w_i) - \mathbf{F}) \cdot \mathbf{n} \, d\ell\right| > \epsilon \mid \mathbf{F} \text{ cont. on } C_a(w_i)\right). \end{aligned}$$

Since  $\rho(n) \in o(1)$ , it holds that  $C_a(w_i) \xrightarrow{P} \{\mathbf{v}\}$ . Moreover,  $\mathbf{F}$  is continuous at  $\mathbf{v}$ , yielding

$$\mathbb{P}(\mathbf{F} \text{ not continuous at some point on } C_a(w_i)) \rightarrow 0.$$

For the second term on the right-hand side of the previous inequality, due to the given continuity of  $\mathbf{F}$  on the arc  $C_a(w_i)$  and the fact that the length of this arc  $d(v, w_i) \Delta\phi(w_i) \xrightarrow{P} 0$  and  $\mathbf{w}_i \in C_a(w_i)$ , it follows that

$$\mathbb{P}\left(\left|\int_{C_a(w_i)} (\mathbf{F}(w_i) - \mathbf{F}) \cdot \mathbf{n} \, d\ell\right| > \epsilon \mid \mathbf{F} \text{ cont. on } C_a(w_i)\right) \rightarrow 0.$$

Consequently, the following convergence result is obtained:

$$\int_{C_a(w_i)} (\mathbf{F}(w_i) - \mathbf{F}) \cdot \mathbf{n} \, d\ell \xrightarrow{P} 0. \quad (39)$$

We combine (38) and (39) using the sum property of convergence in probability into

$$\int_{C_a(w_i)} (\mathbf{F}_g(w_i) - \mathbf{F}) \cdot \mathbf{n} \, d\ell = I_a(w_i) - \int_{C_a(w_i)} \mathbf{F} \cdot \mathbf{n} \, d\ell \xrightarrow{P} 0. \quad (40)$$

The next step is to show that

$$\frac{\mathbf{F}_g(w_i) + \mathbf{F}_g(w_{i+1})}{\|\mathbf{F}_g(w_i) + \mathbf{F}_g(w_{i+1})\|} - \frac{\mathbf{F}(w_i) + \mathbf{F}(w_{i+1})}{\|\mathbf{F}(w_i) + \mathbf{F}(w_{i+1})\|} \xrightarrow{P} \mathbf{0}. \quad (41)$$

Again, we employ the law of total probability to arrive at the following inequality for every  $\epsilon > 0$ :

$$\begin{aligned} & P \left( \left\| \frac{\mathbf{F}_g(w_i) + \mathbf{F}_g(w_{i+1})}{\|\mathbf{F}_g(w_i) + \mathbf{F}_g(w_{i+1})\|} - \frac{\mathbf{F}(w_i) + \mathbf{F}(w_{i+1})}{\|\mathbf{F}(w_i) + \mathbf{F}(w_{i+1})\|} \right\| > \epsilon \right) \\ & \leq P(\mathbf{F}(w_i) + \mathbf{F}(w_{i+1}) = \mathbf{0}) \\ & + P(\mathbf{F}_g(w_i) + \mathbf{F}_g(w_{i+1}) = \mathbf{0} \mid \mathbf{F}(w_i) + \mathbf{F}(w_{i+1}) \neq \mathbf{0}) \\ & + P \left( \left\| \frac{\mathbf{F}_g(w_i) + \mathbf{F}_g(w_{i+1})}{\|\mathbf{F}_g(w_i) + \mathbf{F}_g(w_{i+1})\|} - \frac{\mathbf{F}(w_i) + \mathbf{F}(w_{i+1})}{\|\mathbf{F}(w_i) + \mathbf{F}(w_{i+1})\|} \right\| > \epsilon \right. \\ & \quad \left. \mid \mathbf{F}(w_i) + \mathbf{F}(w_{i+1}) \neq \mathbf{0} \text{ and } \mathbf{F}_g(w_i) + \mathbf{F}_g(w_{i+1}) \neq \mathbf{0} \right). \end{aligned}$$

Similarly to previous parts of the proof, it will be shown that each of the terms on the right-hand side of the above inequality converges to 0. As  $\mathbf{w}_i \xrightarrow{P} \mathbf{v}$  and  $\mathbf{w}_{i+1} \xrightarrow{P} \mathbf{v}$ , the continuity of  $\mathbf{F}$  at  $\mathbf{v}$  yields  $\mathbf{F}(w_i) \xrightarrow{P} \mathbf{F}(\mathbf{v})$  and  $\mathbf{F}(w_{i+1}) \xrightarrow{P} \mathbf{F}(\mathbf{v})$ . The sum property of convergence in probability then implies that

$$\mathbf{F}(w_i) + \mathbf{F}(w_{i+1}) \xrightarrow{P} 2\mathbf{F}(\mathbf{v}).$$

Since  $\|\mathbf{F}(\mathbf{v})\| = 1$ , we get

$$\begin{aligned} & P(\mathbf{F}(w_i) + \mathbf{F}(w_{i+1}) = \mathbf{0}) \\ & \leq P(\|\mathbf{F}(w_i) + \mathbf{F}(w_{i+1})\| < \|\mathbf{F}(\mathbf{v})\|) \\ & = P(2\|\mathbf{F}(\mathbf{v})\| - \|\mathbf{F}(w_i) + \mathbf{F}(w_{i+1})\| > \|\mathbf{F}(\mathbf{v})\|) \\ & \leq P(\|\mathbf{F}(w_i) + \mathbf{F}(w_{i+1}) - 2\mathbf{F}(\mathbf{v})\| > \|\mathbf{F}(\mathbf{v})\|) \rightarrow 0. \end{aligned}$$

In the second term, it is given that  $\|\mathbf{F}(w_i) + \mathbf{F}(w_{i+1})\| > 0$ . Furthermore, (37) holds both for  $w_i$  and  $w_{i+1}$ , which implies that

$$\mathbf{F}_g(w_i) + \mathbf{F}_g(w_{i+1}) - \mathbf{F}(w_i) - \mathbf{F}(w_{i+1}) \xrightarrow{P} \mathbf{0}.$$

Thus, the conditional probability

$$\begin{aligned} & P(\mathbf{F}_g(w_i) + \mathbf{F}_g(w_{i+1}) = \mathbf{0}) \\ & \leq P \left( \|\mathbf{F}_g(w_i) + \mathbf{F}_g(w_{i+1})\| < \frac{\|\mathbf{F}(w_i) + \mathbf{F}(w_{i+1})\|}{2} \right) \\ & \leq P \left( \left\| \mathbf{F}_g(w_i) + \mathbf{F}_g(w_{i+1}) - \mathbf{F}(w_i) - \mathbf{F}(w_{i+1}) \right\| \right. \\ & \quad \left. > \frac{\|\mathbf{F}(w_i) + \mathbf{F}(w_{i+1})\|}{2} \right) \rightarrow 0. \end{aligned}$$

In the third term, it is given that  $\|\mathbf{F}_g(w_i) + \mathbf{F}_g(w_{i+1})\| \geq \delta > 0$  for some  $\delta$ . As a result, exactly the same steps as in

the proof of (37) can be used to bound for every  $\epsilon > 0$  the conditional probability

$$\begin{aligned} & P \left( \left\| \frac{\mathbf{F}_g(w_i) + \mathbf{F}_g(w_{i+1})}{\|\mathbf{F}_g(w_i) + \mathbf{F}_g(w_{i+1})\|} - \frac{\mathbf{F}(w_i) + \mathbf{F}(w_{i+1})}{\|\mathbf{F}(w_i) + \mathbf{F}(w_{i+1})\|} \right\| > \epsilon \right) \\ & \leq P \left( \frac{2\|\mathbf{F}_g(w_i) + \mathbf{F}_g(w_{i+1}) - \mathbf{F}(w_i) - \mathbf{F}(w_{i+1})\|}{\delta} > \epsilon \right) \\ & = P \left( \|\mathbf{F}_g(w_i) + \mathbf{F}_g(w_{i+1}) - \mathbf{F}(w_i) - \mathbf{F}(w_{i+1})\| > \frac{\epsilon\delta}{2} \right) \\ & \rightarrow 0. \end{aligned}$$

Consequently, all three terms converge to 0, which means that their sum also converges to 0, leading to (41).

Afterwards, we define the start and end points of the line segment  $C_l(w_i)$ ,  $\mathbf{s}_i$  and  $\mathbf{t}_i$ , as shown in Fig. 5, and prove that

$$\frac{\mathbf{F}(w_i) + \mathbf{F}(w_{i+1})}{\|\mathbf{F}(w_i) + \mathbf{F}(w_{i+1})\|} - \frac{\mathbf{F}(\mathbf{s}_i) + \mathbf{F}(\mathbf{t}_i)}{\|\mathbf{F}(\mathbf{s}_i) + \mathbf{F}(\mathbf{t}_i)\|} \xrightarrow{P} \mathbf{0}. \quad (42)$$

An intermediate result for the proof of (42) is

$$\mathbf{F}(w_i) + \mathbf{F}(w_{i+1}) - \mathbf{F}(\mathbf{s}_i) - \mathbf{F}(\mathbf{t}_i) \xrightarrow{P} \mathbf{0}. \quad (43)$$

To show this, we start from the following application of the law of total probability for every  $\epsilon > 0$ :

$$\begin{aligned} & P(\|\mathbf{F}(w_i) - \mathbf{F}(\mathbf{s}_i)\| > \epsilon) \\ & \leq P(\nabla u(w_i) = \mathbf{0}) + P(\nabla u(\mathbf{s}_i) = \mathbf{0} \mid \nabla u(w_i) \neq \mathbf{0}) \\ & + P(\|\mathbf{F}(w_i) - \mathbf{F}(\mathbf{s}_i)\| > \epsilon \mid \nabla u(w_i) \neq \mathbf{0} \text{ and } \nabla u(\mathbf{s}_i) \neq \mathbf{0}). \end{aligned}$$

It has already been shown that  $P(\nabla u(w_i) = \mathbf{0}) \rightarrow 0$ . For the second term on the right-hand side,  $\mathbf{w}_i - \mathbf{s}_i \xrightarrow{P} \mathbf{0}$ , since  $\|\mathbf{w}_i - \mathbf{s}_i\| \leq d(v, w_i)\Delta\phi(w_i) \xrightarrow{P} 0$ . Taking into account the fact that  $\mathbf{w}_i \xrightarrow{P} \mathbf{v}$ , we get  $\mathbf{s}_i \xrightarrow{P} \mathbf{v}$ . The continuity of  $\nabla u$  yields  $\nabla u(\mathbf{s}_i) \xrightarrow{P} \nabla u(\mathbf{v})$ . Therefore, we can bound the second term as follows:

$$\begin{aligned} & P(\nabla u(\mathbf{s}_i) = \mathbf{0}) \leq P \left( \|\nabla u(\mathbf{s}_i)\| < \frac{\|\nabla u(\mathbf{v})\|}{2} \right) \\ & \leq P \left( \|\nabla u(\mathbf{s}_i) - \nabla u(\mathbf{v})\| > \frac{\|\nabla u(\mathbf{v})\|}{2} \right) \rightarrow 0. \end{aligned}$$

In the third term, it is given that  $\|\nabla u(w_i)\| \geq \delta > 0$  for some  $\delta$ . We thus follow the same steps as in the proof of (37) to bound the respective conditional probability:

$$\begin{aligned} & P(\|\mathbf{F}(w_i) - \mathbf{F}(\mathbf{s}_i)\| > \epsilon) \\ & \leq P \left( \frac{2\|\nabla u(w_i) - \nabla u(\mathbf{s}_i)\|}{\delta} > \epsilon \right) \\ & = P \left( \|\nabla u(w_i) - \nabla u(\mathbf{s}_i)\| > \frac{\epsilon\delta}{2} \right) \rightarrow 0 \quad \forall \epsilon > 0, \end{aligned}$$

where the limit in the last line holds due to the fact that  $\nabla u(w_i) - \nabla u(\mathbf{s}_i) \xrightarrow{P} \mathbf{0}$ . As a result, it holds that  $\mathbf{F}(w_i) - \mathbf{F}(\mathbf{s}_i) \xrightarrow{P} \mathbf{0}$  and in a similar fashion  $\mathbf{F}(w_{i+1}) - \mathbf{F}(\mathbf{t}_i) \xrightarrow{P} \mathbf{0}$ . The sum property of convergence in probability directly implies (43) from the above two results.

To prove (42), we use the law of total probability to write

$$\begin{aligned} & P \left( \left\| \frac{\mathbf{F}(w_i) + \mathbf{F}(w_{i+1})}{\|\mathbf{F}(w_i) + \mathbf{F}(w_{i+1})\|} - \frac{\mathbf{F}(\mathbf{s}_i) + \mathbf{F}(\mathbf{t}_i)}{\|\mathbf{F}(\mathbf{s}_i) + \mathbf{F}(\mathbf{t}_i)\|} \right\| > \epsilon \right) \\ & \leq P(\mathbf{F}(w_i) + \mathbf{F}(w_{i+1}) = \mathbf{0}) \\ & + P(\mathbf{F}(\mathbf{s}_i) + \mathbf{F}(\mathbf{t}_i) = \mathbf{0} | \mathbf{F}(w_i) + \mathbf{F}(w_{i+1}) \neq \mathbf{0}) \\ & + P \left( \left\| \frac{\mathbf{F}(w_i) + \mathbf{F}(w_{i+1})}{\|\mathbf{F}(w_i) + \mathbf{F}(w_{i+1})\|} - \frac{\mathbf{F}(\mathbf{s}_i) + \mathbf{F}(\mathbf{t}_i)}{\|\mathbf{F}(\mathbf{s}_i) + \mathbf{F}(\mathbf{t}_i)\|} \right\| > \epsilon \right. \\ & \quad \left. | \mathbf{F}(w_i) + \mathbf{F}(w_{i+1}) \neq \mathbf{0} \text{ and } \mathbf{F}(\mathbf{s}_i) + \mathbf{F}(\mathbf{t}_i) \neq \mathbf{0} \right). \quad (44) \end{aligned}$$

We have already proved

$$P(\mathbf{F}(w_i) + \mathbf{F}(w_{i+1}) = \mathbf{0}) \rightarrow 0.$$

For the second term on the right-hand side, it is given that  $\|\mathbf{F}(w_i) + \mathbf{F}(w_{i+1})\| > 0$ . We bound the conditional probability

$$\begin{aligned} & P(\mathbf{F}(\mathbf{s}_i) + \mathbf{F}(\mathbf{t}_i) = \mathbf{0}) \\ & \leq P \left( \|\mathbf{F}(\mathbf{s}_i) + \mathbf{F}(\mathbf{t}_i)\| < \frac{\|\mathbf{F}(w_i) + \mathbf{F}(w_{i+1})\|}{2} \right) \\ & \leq P \left( \|\mathbf{F}(w_i) + \mathbf{F}(w_{i+1}) - \mathbf{F}(\mathbf{s}_i) - \mathbf{F}(\mathbf{t}_i)\| \right. \\ & \quad \left. > \frac{\|\mathbf{F}(w_i) + \mathbf{F}(w_{i+1})\|}{2} \right) \rightarrow 0, \end{aligned}$$

where the last limit is due to (43). Last, we focus on the third term, where it is given that  $\|\mathbf{F}(w_i) + \mathbf{F}(w_{i+1})\| \geq \delta > 0$  for some  $\delta$ . Taking the same steps as in the proof of (37) to bound the third term, we obtain

$$\begin{aligned} & P \left( \left\| \frac{\mathbf{F}(w_i) + \mathbf{F}(w_{i+1})}{\|\mathbf{F}(w_i) + \mathbf{F}(w_{i+1})\|} - \frac{\mathbf{F}(\mathbf{s}_i) + \mathbf{F}(\mathbf{t}_i)}{\|\mathbf{F}(\mathbf{s}_i) + \mathbf{F}(\mathbf{t}_i)\|} \right\| > \epsilon \right) \\ & \leq P \left( \frac{2\|\mathbf{F}(w_i) + \mathbf{F}(w_{i+1}) - \mathbf{F}(\mathbf{s}_i) - \mathbf{F}(\mathbf{t}_i)\|}{\delta} > \epsilon \right) \\ & = P \left( \|\mathbf{F}(w_i) + \mathbf{F}(w_{i+1}) - \mathbf{F}(\mathbf{s}_i) - \mathbf{F}(\mathbf{t}_i)\| > \frac{\epsilon\delta}{2} \right) \\ & \rightarrow 0 \quad \forall \epsilon > 0. \end{aligned}$$

The sum of all three terms on the right-hand side of (44) converges to 0 as each of them converges to 0, which implies (42).

For the line segment  $C_l(w_i)$ , we define the event

$$B = \exists \mathbf{r} \in C_l(w_i) : \mathbf{F}(\mathbf{r}) = \frac{\mathbf{F}(\mathbf{s}_i) + \mathbf{F}(\mathbf{t}_i)}{\|\mathbf{F}(\mathbf{s}_i) + \mathbf{F}(\mathbf{t}_i)\|}$$

and show that

$$P(B) \rightarrow 1. \quad (45)$$

Let us consider the complement of  $B$ . The law of total probability implies that

$$\begin{aligned} P(B^c) & \leq P(\mathbf{F} \text{ not continuous at some point on } C_l(w_i)) \\ & + P(B^c | \mathbf{F} \text{ continuous on } C_l(w_i)). \end{aligned}$$

Since  $C_l(w_i) \xrightarrow{P} \{\mathbf{v}\}$  and  $\mathbf{F}$  is continuous at  $\mathbf{v}$ , it holds for the first term on the right-hand side that

$$P(\mathbf{F} \text{ not continuous at some point on } C_l(w_i)) \rightarrow 0.$$

In the second term, it is given that  $\mathbf{F}$  is continuous on  $C_l(w_i)$ . We note that  $\mathbf{F}$  is a circular quantity, since  $\|\mathbf{F}\| = 1$ . Therefore, if we define the event

$$D = \exists \mathbf{q} \in C_l(w_i) : \mathbf{F}(\mathbf{q}) = -\frac{\mathbf{F}(\mathbf{s}_i) + \mathbf{F}(\mathbf{t}_i)}{\|\mathbf{F}(\mathbf{s}_i) + \mathbf{F}(\mathbf{t}_i)\|},$$

the intermediate value theorem for  $\mathbf{F}$  on  $C_l(w_i)$  implies that  $B \cup D = \Omega$ . It is straightforward that

$$\left\| -\frac{\mathbf{F}(\mathbf{s}_i) + \mathbf{F}(\mathbf{t}_i)}{\|\mathbf{F}(\mathbf{s}_i) + \mathbf{F}(\mathbf{t}_i)\|} - \mathbf{F}(\mathbf{s}_i) \right\| \geq \sqrt{2}.$$

However, since it holds for every  $\mathbf{p} \in C_l(w_i)$  that  $\|\mathbf{p} - \mathbf{v}\| \leq \rho(n)$ , we have  $\mathbf{p} \xrightarrow{P} \mathbf{v}$ . It was previously shown that  $\mathbf{s}_i \xrightarrow{P} \mathbf{v}$ , so the continuity of  $\mathbf{F}$  at  $\mathbf{v}$  yields  $\mathbf{F}(\mathbf{p}) \xrightarrow{P} \mathbf{F}(\mathbf{v})$  and  $\mathbf{F}(\mathbf{s}_i) \xrightarrow{P} \mathbf{F}(\mathbf{v})$ . It follows that

$$P(\|\mathbf{F}(\mathbf{p}) - \mathbf{F}(\mathbf{s}_i)\| \geq \epsilon) \rightarrow 0 \quad \forall \mathbf{p} \in C_l(w_i) \quad \forall \epsilon > 0.$$

If we consider the point  $\mathbf{q}$  which appears in the expression of event  $D$ , we obtain

$$P(D) \leq P(\exists \mathbf{q} \in C_l(w_i) : \|\mathbf{F}(\mathbf{q}) - \mathbf{F}(\mathbf{s}_i)\| \geq \sqrt{2}) \rightarrow 0.$$

As a result, when it is given that  $\mathbf{F}$  is continuous on  $C_l(w_i)$ , it holds that

$$P(B^c) = 1 - P(B) \leq P(D) \rightarrow 0.$$

Since both terms of the sum that bounds  $P(B^c)$  converge to 0, so does  $P(B^c)$ , leading to (45). Thus, there exists a sequence of random variables  $\{\mathbf{r}_n\}$  such that every term of the sequence belongs to  $(C_l(w_i))_n$  and

$$\frac{\mathbf{F}(\mathbf{s}_i) + \mathbf{F}(\mathbf{t}_i)}{\|\mathbf{F}(\mathbf{s}_i) + \mathbf{F}(\mathbf{t}_i)\|} - \mathbf{F}(\mathbf{r}_n) \xrightarrow{P} \mathbf{0}. \quad (46)$$

For simplicity, we omit the index  $n$  from  $\mathbf{r}_n$  and write  $\mathbf{r}$  instead, as we have done for all the rest of the sequences of random variables that are treated in the proof. For this  $\mathbf{r}$ , we combine (41), (42) and (46) through summation into

$$\frac{\mathbf{F}_g(w_i) + \mathbf{F}_g(w_{i+1})}{\|\mathbf{F}_g(w_i) + \mathbf{F}_g(w_{i+1})\|} - \mathbf{F}(\mathbf{r}) \xrightarrow{P} \mathbf{0}. \quad (47)$$

Taking into account the closed form (22) for  $I_l(w_i)$  and the product and sum properties of convergence in probability, we get

$$\int_{C_l(w_i)} \left( \frac{\mathbf{F}_g(w_i) + \mathbf{F}_g(w_{i+1})}{\|\mathbf{F}_g(w_i) + \mathbf{F}_g(w_{i+1})\|} - \mathbf{F}(\mathbf{r}) \right) \cdot \mathbf{n} \, d\ell \xrightarrow{P} 0. \quad (48)$$

Additionally, the law of total probability yields for every  $\epsilon > 0$

$$\begin{aligned} & P \left( \left| \int_{C_l(w_i)} (\mathbf{F}(\mathbf{r}) - \mathbf{F}) \cdot \mathbf{n} \, d\ell \right| > \epsilon \right) \\ & \leq P(\mathbf{F} \text{ not continuous at some point on } C_l(w_i)) \\ & + P \left( \left| \int_{C_l(w_i)} (\mathbf{F}(\mathbf{r}) - \mathbf{F}) \cdot \mathbf{n} \, d\ell \right| > \epsilon | \mathbf{F} \text{ cont. on } C_l(w_i) \right). \end{aligned}$$

It has been shown that the first term on the right-hand side converges to 0. The second term, which is conditioned on  $\mathbf{F}$

being continuous on  $C_l(w_i)$ , also converges to 0, since the length of  $C_l(w_i)$ , which is equal to  $|d(v, w_{i+1}) - d(v, w_i)|$ , converges to 0 almost surely. Consequently, we obtain

$$\int_{C_l(w_i)} (\mathbf{F}(\mathbf{r}) - \mathbf{F}) \cdot \mathbf{n} d\ell \xrightarrow{P} 0. \quad (49)$$

The summation of (48) and (49) leads to

$$I_l(w_i) - \int_{C_l(w_i)} \mathbf{F} \cdot \mathbf{n} d\ell \xrightarrow{P} 0. \quad (50)$$

Due to the fact that (40) and (50) hold for every  $i \in \{1, \dots, N\}$ , the sum property of convergence in probability implies (24).

Finally, our construction of region  $S(v)$  guarantees that

$$S(v) \xrightarrow{\text{a.s.}} \{\mathbf{v}\}$$

because  $\rho(n) \in o(1)$ . Thus, it holds

$$\frac{\oint_{\Gamma(S(v))} \mathbf{F} \cdot \mathbf{n} d\ell}{|S(v)|} \xrightarrow{\text{a.s.}} \kappa(v).$$

The last result, together with (24), implies that our approximation (23) converges in probability to the true value of the curvature, after making use of the triangle inequality. ■

## REFERENCES

- [1] M. Kass, A. Witkin, and D. Terzopoulos, “Snakes: Active contour models,” *International Journal of Computer Vision*, vol. 1, no. 4, pp. 321–331, 1988.
- [2] V. Caselles, F. Catté, T. Coll, and F. Dibos, “A geometric model for active contours in image processing,” *Numerische Mathematik*, vol. 66, pp. 1–31, 1993.
- [3] V. Caselles, R. Kimmel, and G. Sapiro, “Geodesic active contours,” *International Journal of Computer Vision*, vol. 22, no. 1, pp. 61–79, 1997.
- [4] S. Osher and J. A. Sethian, “Fronts propagating with curvature dependent speed: Algorithms based on Hamilton-Jacobi formulations,” *Journal of Computational Physics*, vol. 79, pp. 12–49, 1988.
- [5] L. Vincent, “Graphs and mathematical morphology,” *Signal Processing*, vol. 16, pp. 365–388, 1989.
- [6] J. Jaromczyk and G. Toussaint, “Relative neighborhood graphs and their relatives,” *Proceedings of the IEEE*, vol. 80, no. 9, pp. 1502–1517, Sep. 1992.
- [7] H. J. A. M. Heijmans, P. Nacken, A. Toet, and L. Vincent, “Graph morphology,” *Journal of Visual Communication and Image Representation*, vol. 3, no. 1, pp. 24–38, Mar. 1992.
- [8] J. Cousty, L. Najman, and J. Serra, “Some morphological operators in graph spaces,” in *Proc. International Symposium on Mathematical Morphology*, 2009.
- [9] V.-T. Ta, A. Elmoataz, and O. Lézoray, “Nonlocal PDEs-based morphology on weighted graphs for image and data processing,” *IEEE Trans. Image Process.*, vol. 20, no. 6, pp. 1504–1516, Jun. 2011.
- [10] O. Lézoray, A. Elmoataz, and V.-T. Ta, “Nonlocal PDEs on graphs for active contours models with applications to image segmentation and data clustering,” in *ICASSP*, 2012, pp. 873–876.
- [11] F. Lozes, A. Elmoataz, and O. Lézoray, “Partial difference operators on weighted graphs for image processing on surfaces and point clouds,” *IEEE Trans. Image Process.*, vol. 23, no. 9, pp. 3896–3909, Sep. 2014.
- [12] K. Drakopoulos and P. Maragos, “Active contours on graphs: Multiscale morphology and graphcuts,” *IEEE J. Sel. Topics Signal Process.*, vol. 6, no. 7, pp. 780–794, Nov. 2012.
- [13] J. Shi and J. Malik, “Normalized cuts and image segmentation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, Aug. 2000.
- [14] Y. Boykov, O. Veksler, and R. Zabih, “Fast approximate energy minimization via graph cuts,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 11, pp. 1222–1239, Nov. 2001.
- [15] Y. Boykov and V. Kolmogorov, “Computing geodesics and minimal surfaces via graph cuts,” in *Proc. Int. Conf. Comput. Vis.*, vol. 1, Oct. 2003, pp. 26–33.
- [16] J. Cousty, G. Bertrand, L. Najman, and M. Couprie, “Watershed cuts: Minimum spanning forests and the drop of water principle,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 8, pp. 1362–1374, Aug. 2009.
- [17] —, “Watershed cuts: Thinnings, shortest path forests, and topological watersheds,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 5, pp. 925–939, May 2010.
- [18] C. Couprie, L. Grady, L. Najman, and H. Talbot, “Power watershed: A unifying graph-based optimization framework,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 7, pp. 1384–1399, Jul. 2011.
- [19] D. Martin, C. Fowlkes, D. Tal, and J. Malik, “A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics,” in *Proc. Int. Conf. Comput. Vis.*, vol. 2, Jul. 2001, pp. 416–423.